Probabilistic Choice over a Continuous Range: An Econometric Model Based on Extreme-Value Stochastic Processes

Stephen R. Cosslett[1]

Department of Economics
Ohio State University

(Revised September 1988)

Abstract

The behaviour of an individual whose choice set consists of a real interval $[0, T]$ (for example, time or location) can be modelled in terms of random utility maximization, like probabilistic models of discrete choice. The chosen point t maximizes a utility function defined over the choice set. The utility function has a deterministic part which is a function of observed variables and unknown parameters to be estimated, and a stochastic part which is known to the decision-maker but not to the econometrician. Specifying a probability distribution for the stochastic part then induces a probability distribution of the choice t over the interval $[0, T]$. The resulting choice probabilities will depend on observed variables and on unknown parameters, which can then be estimated from data on observed choices. In the class of models proposed here, the stochastic utility is represented by a continuous-time random process whose finite-dimensional distributions are based on the generalized extreme-value distribution.

First some general results are presented, showing how the distribution functions of the stochastic process determine the distribution function of maximum utility, which in turn determines the probability that utility is maximized in a given interval of the choice set. Then we investigate the properties of a particular class of extreme-value stochastic processes, which are shown to have continuous sample paths with probability one. This makes them plausible candidates for the random part of a utility function (in contrast with, for example, a white-noise process). Results are established for the distribution of the utility maximum. Under suitable regularity conditions on the systematic part of the utility function, an algorithm leads to

[1] I would like to thank D. McFadden for suggesting the problem of maximizing a continuous-time stochastic utility function. This research was supported in part by the Center for Econometrics and Decision Sciences at the University of Florida.

computationally tractable, but not simple, expressions for the choice probabilities. These choice probabilities are shown to generate a probability measure on $[0,T]$.

Some special cases have to be addressed: points with discrete choice probabilities, intervals with zero choice probability, and additional regularity conditions that ensure the existence of a choice probability density function. Following the formal results, we give explicit expressions for the choice probabilities in some simple cases, to illustrate the effects of peaks, valleys, and steps in the systematic part of the utility function. Finally, we discuss the question of parameter estimation, and give some conditions for consistent maximum likelihood estimation in the case of discretized data on the choice variable.

Keywords: choice probability model, continuous choice, generalized extreme value, random utility

Address: Stephen R. Cosslett, Department of Economics, Ohio State University, Columbus, OH 43210.

1. Introduction

This paper presents a model of probabilistic choice over a continuous range of alternatives, based on maximization of an underlying utility function. The individual who makes the choice knows the utility function, and so his choice is entirely deterministic. The econometrician, on the other hand, cannot observe the utility function, and so models it as a stochastic process over the choice set, in terms of observed variables and unknown parameters. A model of the stochastic utility function then induces a probability distribution for the location of its maximum. This probability distribution is the choice probability model. If it is not too complicated, its parameters can then be estimated from data on observed choices.

As an illustration, suppose that an individual has to schedule an activity to start at time t in an interval $[0,T]$. There is an observable vector function $x(t)$ that expresses the attributes of the activity as functions of t. For example, if the activity is an automobile trip and t is departure time, then one component of $x(t)$ might be the time taken to reach the destination. There is also an observed vector z of characteristics of the individual, which in principle help to determine the structure of his utility function. Since there will always be other relevant variables that are not observed, the choice t^* cannot be uniquely determined from $x(\cdot)$ and z. Instead, the best we can do is to model the choice probabilities $p(t) = p(t \mid x, z)$, where $p(t)\,dt$ is the probability that t^* will be in the small interval $[t, t+dt]$.

This is obviously very much like discrete choice modelling, the difference being the continuous choice set $[0,T]$ instead of choice from a finite set of alternatives. A large class of discrete choice probability models has been developed based on random utility maximization (see, for example, McFadden, 1984). Each individual chooses among the alternatives as if he were maximizing some objective function, which we may call utility. Individual values of this function are modelled as random vectors drawn according to some probability distribution, conditional on observed variables and on an unknown set of parameters. That leads to a choice probability model: the probability that an alternative is chosen is just the probability that its component of the random utility vector is greater than that of any other alternative.

Several considerations go into the specification of the underlying probability distribution. Among others, we would like compatibility with the requirements of economic theory; inherent plausibility of the structure of the model; enough flexibility that we are not led to a contradiction of observed data; and computational tractability of the choice probabilities and of the parameter estimates. Computational tractability is obviously important, and will be one of the concerns of this paper. Nevertheless, it is generally held that a choice probability model ought to be compatible with some underlying utility maximization model, as opposed to a purely ad hoc set of probabilities, and it is from that point of view that we propose a model of choice over a continuous range. Of course there is always some stochastic utility function corresponding to any given choice probability model $p(t)$: for example, $U(t) = 1\{t = t^*\}$ where t^* is a random variable with probability density function p. We shall therefore have to discuss what properties a stochastic process ought to have in order to be a reasonable model for a stochastic utility function.

A discrete choice model is based on the distribution of a random utility vector, with dimension equal to the number of alternatives in the choice set. For continuous choice, we have to specify instead the distribution of a stochastic process on $[0,T]$. If one starts from some standard model such as a Gaussian Markov process for the stochastic utility, one soon runs into severe difficulties because there is no analytic expression for the probability distribution of the maximum, except in some very special cases.

A different line of attack is to study the behavior of discrete choice models as the number of alternatives becomes large. Ben-Akiva and Watanatada (1981), among others, proposed the continuous logit model for spatial choice: it is obtained from the conventional multinomial logit model by allowing the number of alternatives to tend to infinity, sums being replaced by the corresponding integrals. Another approach is that taken by Small (1987): the range of choices is partitioned into a number of intervals, each of which can then be considered as a discrete alternative. The resulting discrete choice model is then represented by Small's ordered generalized extreme value (OGEV) model, which is a particular case of the generalized extreme value model introduced by McFadden (1978, 1981). This allows for a realistic pattern of serial correlation between the random utilities corresponding to the sequence of the intervals, in contrast to the multinomial logit model. These previous studies, however, did not address the question of whether the choice probabilities are compatible with maximization of some reasonable underlying stochastic process.

The model presented here is related to both of these last two approaches. It is based on the GEV model, extended to a continuous range of alternatives, and it leads to computationally tractable (although not always simple) choice probabilities.[2] Explicit expressions for the choice probabilities are worked out in some special cases as an illustration. Somewhat surprisingly, the continuous logit model can emerge as a special case if the attribute vector $x(t)$ varies sufficiently slowly over time.

First we investigate the general problem of defining the choice probability function for a class of stochastic utility processes, subject to some continuity conditions. We propose the following restriction on the process, if we want to plausibly interpret its realisations as utility functions: with probability one, sample paths should be continuous, except for a (fixed) finite set of points where there may be jumps. More general models are of course possible: one could, for example, allow the set of discontinuity points to be random also, although it is not known whether a tractable model of this kind can be found. Further regularity conditions could also be proposed, which might not hold for the model as presently formulated: for example, although the finite-dimensional distribution functions of the stochastic process are continuous, they do not correspond to absolutely continuous probability measures. Further research should show what additional regularity conditions can be imposed on models of this type.

General results are presented in Section 2, for the case where the utility function is the sum of an unobserved stochastic process and a parametric function of observed variables. The distribution function of maximum utility, G, is related to the finite-dimensional distributions of the stochastic process, and the probability of the event that utility is maximized in some interval J of the choice set, $P_0(J)$, is expressed in terms of G.

In Section 3, we introduce a specific stochastic process based on the GEV model. Some basic properties of the process are derived, including continuity in probability. Using a result of de Haan (1984) on a spectral representation of max-stable processes, we then find the almost sure continuity of sample paths referred to above. Next we apply the general results of Section 2 to the GEV model of Section 3. In Section 4 we derive an expression for the choice probabilities, defined on intervals of the choice set, in terms of the parametric part of the utility function. We also show that these choice probabilities can be extended to a probability measure. In Section 5 we consider some special cases of the choice probability function: steps or sharp peaks in the parametric utility function can lead to discrete points which have positive choice probability; there is the possibility of "shadows", i.e., intervals for which the choice probability is zero; and finally, under

---

[2] A recent paper by Dagsvik (1988) also extends the generalized extreme value model to a continuous choice set.

suitable assumptions, there is the "regular" case where a choice probability density function can be defined. To illustrate the construction of the choice probabilities, some simple examples are worked out explicitly in Section 6. Maximum likelihood estimation of the parameters is discussed briefly in Section 7, for the case where observations have been categorized into discrete intervals. The Appendix contains proofs of the formal results, together with some preliminary lemmas.

Finally, we should emphasize that the model presented here is not meant to be definitive; rather, it demonstrates an approach that may lead to a variety of more general models of stochastic utility maximization.

2. Choice probabilities and stochastic utility in continuous time

Let $U(t,z)$ be the conditional indirect utility of someone with characteristics z, conditioned on $t \in [0,T]$, where t is the starting time of the activity. That is, $U(t,z)$ is the utility attained by optimizing over all other decisions while keeping t fixed. In this paper we consider only the additively separable case

$$U(t,z) = V(x(t), z, \theta) + W(t) \tag{2.1}$$

where V is a known parametric function of $x(t)$ (the observed characteristics of the activity if started at time t) and of z, with unknown parameter vector θ. The other term, $W(\cdot)$, is an unobserved stochastic process. The distribution of W is supposed to be known up to a parameter vector ϕ; it may also depend on z but we assume it does not depend on $x(\cdot)$.

The non-stochastic part $V(t,\theta) = V(x(t), z, \theta)$ is assumed to satisfy the following regularity conditions.

(V.1) $V(t,\theta)$ is continuous in t, except possibly at a finite set of points that does not depend on x, z or θ.

(V.2) At points of discontinuity in t, the left and right limits $V(t-,\theta)$ and $V(t+,\theta)$ exist, and $V(t,\theta) = \max\{V(t-,\theta), V(t+,\theta)\}$.

(V.3) $\partial V(t,\theta)/\partial t$ is continuous in t, except possibly at a finite set of points that does not depend on x, z or θ.

(V.4) $V(t,\theta)$ is continuous in θ uniformly in t.

Conditions V.1 and V.3 allow, for example, a piecewise linear parametrization of V. There can also be jumps in V at fixed times, which could represent, for example, time-varying user fees. Condition V.2 implies that *V* is upper semicontinuous in *t*. Without this last property, the utility function might not have a maximum and the choice problem would be ill-defined. Condition V.4 is not needed for formulating the choice probability model, but will be used later when we come to estimate θ.

The finite-dimensional distributions of *W* are

$$F_{t_1,\ldots,t_k}(w_1,\ldots,w_k) = \Pr\{W(t_i) \le w_i,\ i = 1,\ldots,k\} \tag{2.2}$$

for $t_i \in [0,T]$, $i = 1,\ldots,k$, and $k \in \{1,2,\ldots\}$. Without loss we can arrange that $t_1 < t_2 < \ldots < t_k$. For notational convenience we have not explicitly shown the dependence on ϕ and z. The finite-dimensional distribution functions must satisfy the consistency condition

$$\lim_{w_i \to \infty} F_{t_1,\ldots,t_k}(w_1,\ldots,w_k) = F_{t_1,\ldots,t_{i-1},t_{i+1},\ldots,t_k}(w_1,\ldots,w_{i-1},w_{i+1},\ldots,w_k) \tag{2.3}$$

for each $i \in \{1,\ldots,k\}$ and $k \ge 2$. It then follows that there is a separable stochastic process *W* on $[0,T]$ with those finite-dimensional distributions (see, for example, Billingsley (1986), Section 38). Separable means that there is a countable dense set $D \subset [0,T]$, such that for any $t \in [0,T]$ and for almost any sample path (i.e., except for a set of sample paths that has probability zero) there is a sequence $\{t_1,t_2,\ldots\}$ of points in *D* with

$$t_n \to t \quad \text{and} \quad w(t_n) \to w(t) \tag{2.4}$$

(This is different from continuity, which requires (2.4) to hold for any sequence that converges to *t*). The condition that *W* should be separable still may not define the process uniquely. However, the results that we obtain will apply to any separable stochastic process *W* that has the finite-dimensional distributions of the model.

By itself the existence of a separable stochastic process is not very useful, because it holds for any consistent set of finite-dimensional distributions (including, for example, those corresponding to the independent logit model), and because it says little about the properties of the stochastic process *W*. If *W* is to be a plausible candidate for the unobserved part of the utility function $U(\cdot,z)$, it should satisfy some regularity properties. For example, an undesirable property of models with white-noise type processes, where $W(t_1)$ and $W(t_2)$ are independent for any $t_1 \ne t_2$, is that the utility at the chosen point is infinite. In the extreme-value model for *W* proposed in the next section, we find that the sample paths are continuous with probability one. In particular, this implies that $\max_t\{W(t)\}$ is finite with probability one.

To avoid repetition, we shall stop saying "with probability one": any event E is to be interpreted as $E \cap N^c$, where N^c is some fixed event with probability one. In those results that depend on continuity, N^c will be the event that the sample path $W(t)$ is continuous.

Proposition 2.1. (a) If W is separable, and if V satisfies condition V.1, then U is separable.
(b) If sample paths of W are continuous, and if V satisfies conditions V.1–V.2, then sample paths of U are upper semicontinuous.

The proof is omitted: these results follow directly from the definitions.

Now consider the maximization of $U(t)$. Let

$$M = \{t \mid U(t) \geq U(s) \text{ for all } s \in [0,T]\}. \tag{2.5}$$

If $U(t)$ is upper semicontinuous then M is non-empty, because the choice set is compact. For any interval $J \subset [0,T]$, let

$$P_0(J) = \Pr\{M \subseteq J\}. \tag{2.6}$$

If M is a single point (i.e., if there are no multiple maxima) this is the same as the choice probability

$$P(J) = \Pr\{t^* \in J\}, \tag{2.7}$$

but we should not assume this in advance. To see that the probability (2.6) is well defined, we need the following result.

Lemma 2.1. Suppose U is separable, and its sample paths are upper semicontinuous. Let J be an interval of $[0,T]$. Then $\{M \subseteq J\}$ is an element of $\mathcal{R}^{[0,T]}$, the σ-field generated by cylinders in $\mathrm{R}^{[0,T]}$.

Because the finite-dimensional distributions define a probability measure on $\mathcal{R}^{[0,T]}$, (2.6) defines $P_0(J)$ for each interval $J \subset [0,T]$. Obviously $P_0([0,T]) = 1$ and $P_0(\varnothing) = 0$, but we cannot yet conclude that P_0 extends to a probability measure. For example, it does not follow that $P_0(J_1 \cup J_2) = P_0(J_1) + P_0(J_2)$ for disjoint intervals J_1 and J_2, because there might be points of M in both intervals.

We shall investigate the properties of P_0 when W is defined by the extreme-value model of Section 3. In Section 4, $P_0(J)$ is evaluated for any interval $J \subset [0,T]$. Then we shall find that $P_0(J) + P_0(J^c) = 1$, which means $P_0(J)$ is unambiguously identified with the choice probability for that interval, i.e., $P_0(J) = P(J)$. We shall also verify that

P_0 satisfies the countable additivity property on intervals of $[0,T]$, from which it follows that P_0 extends to a probability measure, i.e., the choice probability $P(A)$ is defined for any Borel set $A \subseteq [0,T]$.

The construction used to prove lemma 2.1 does not lead to a practical method of evaluating $P_0(J)$. The following results show how, at least in principle, $P_0(J)$ can be constructed from limits of finite-dimensional distribution functions. Of course they are useful only if the limits are tractable and satisfy the differentiability condition of Lemma 2.3. First we need the distribution of the maximum utility,

$$\begin{aligned} G(u\,;V) &= \Pr\{\sup_{t\in[0,T]} U(t) \le u\} \\ &= \Pr\{W(t) \le u - V(t) \text{ for all } t \in [0,T]\} \end{aligned} \tag{2.8}$$

Lemma 2.2. Suppose *W* is separable, and condition V.1 holds. Let *J* be an interval of $[0,T]$. Let $J_n = \{t_1,...,t_{m(n)}\}$, $n = 1,2,...$, denote a collection of finite subsets of *J*. Then there is a countable dense set *D* such that if $J_n \uparrow J \cap D$ then

$$\Pr\{\sup_{t\in J} U(t) \le u\} = \lim_{n\to\infty} F_{\{J_n\}}[u - V(t_1),...,u - V(t_{m(n)})] \tag{2.9}$$

where *F* is the finite-dimensional distribution function of *W*.

If in fact *W* is continuous and *V* also satisfies condition V.2, then we can use any countable dense *D* in constructing the sets J_n, for example R_0 (the rational numbers).

Next, let *J* be an open interval with respect to [0, T] and define

$$U_1 = \sup_{t\in J} U(t), \qquad U_2 = \sup_{t\in J^c} U(t), \tag{2.10}$$

where J^c is the complement of *J* with respect to $[0,T]$. Once we have found *G*, the distribution of maximum utility, then we also have the joint distribution of U_1 and U_2, because

$$\begin{aligned} &\Pr\{U_1 \le u_1, U_2 \le u_2\} \\ &= \Pr\{W(t) \le u_1 - V(t) \text{ for all } t \in J,\quad W(t) \le u_2 - V(t) \text{ for all } t \in J^c\} \\ &= \Pr\{W(t) \le u_2 - V(t) + (u_1 - u_2)\,1\{J\} \text{ for all } t\} \\ &= G(u_2\,; V - (u_1 - u_2)\,1\{J\}) \end{aligned} \tag{2.11}$$

where $1\{J\}$ is the indicator function of *J*. But

$$P_0(J) = \Pr\{U_1 > U_2\}. \tag{2.12}$$

so $P_0(J)$ can be found from the joint distribution (2.11) just as in the stochastic utility model of binary choice. The result is in the next lemma.

Lemma 2.3. Let $G(\cdot;V)$ be the distribution function of maximum utility, defined by (2.8), and let *J* be an open interval with respect to $[0,T]$. Suppose there is a continuous function G_1 such that, given any point *u* and any $\delta > 0$,

$$\left| \frac{1}{\varepsilon}\big(G(u';V-\varepsilon 1\{J\}) - G(u';V)\big) - G_1(u') \right| \le \delta \tag{2.13}$$

for all sufficiently small $\varepsilon > 0$ and all u' in some neighborhood of *u*. Then

$$P_0(J) = \int_{-\infty}^{\infty} du\, G_1(u)\,. \tag{2.14}$$

To simplify the notation, the explicit dependence of G_1 on *V* and *J* has been dropped. To get any further, we need a specific model for *W*.

3. The GEV stochastic process and its properties

We now introduce a specific model for $W(t)$, the stochastic part of the utility function (2.1). Define the functions $H_k(y_1,...,y_k;t_1,...,t_k)$ in terms of the finite-dimensional distribution functions (2.2) by

$$F_{t_1,...,t_k}(w_1,...,w_k) = \exp\{-H_k(e^{-w_1},...,e^{-w_k};t_1,...,t_k)\} \tag{3.1}$$

for $k = 1,2,...$ Suppose $t \in A \subset [0,T]$, where *A* is any finite set of points. Then the following conditions on the functions H_k are sufficient for the functions *F*, with $\{t_1,...,t_k\} \subseteq A$, to be a set of probability distribution functions (McFadden, 1978, 1981).

(G.1) $H_k(y_1,...,y_k;t_1,...,t_k)$ is nonnegative and linearly homogeneous in $y_1,...,y_k$ for $y_i \in \mathrm{R}^+$, $i = 1,...,k$.

(G.2) $\lim_{y_i \to \infty} H_k(y_1,...,y_k;t_1,...,t_k) = \infty$ for any $i \in \{1,...,k\}$.

(G.3) If $\{i(1),...,i(r)\} \subseteq \{1,...,k\}$, then the mixed partial derivative $\partial^r H_k / \partial y_{i(1)} ... \partial y_{i(r)}$ exists, and is nonnegative for odd *r* and nonpositive for even *r*.

(G.4) $H_k(y_1,...,y_{i-1},0,y_{i+1},...,y_k;t_1,...,t_k) =$

$H_{k-1}(y_1,...,y_{i-1},y_{i+1},...,y_k;t_1,...,t_{i-1},t_{i+1},...,t_k)$ for $k \ge 2$.

For discrete choice, these conditions define the class of generalized extreme value (GEV) choice probability functions (McFadden, 1978). The linear homogeneity in condition G.1

is not necessary for the functions F to be distribution functions, except in that it implies $H_1(0;t_1)=0$, but it allows one to derive an explicit expression for the choice probabilities. Condition G.3 ensures that there is a nonnegative density function corresponding to the distribution function *F*. Condition G.4 restates the consistency condition (2.3).

In the continuous case $t \in [0,T]$, consider an extreme value stochastic process (EVSP) with finite dimensional distributions given by (3.1) with this H_k:

$$H_k(y_1,...,y_k;t_1,...,t_k) = \int dt \max_i \{ y_i \exp[-\lambda(t_i - t)] \} \tag{3.2}$$

where the maximum is over $i = 1,...,k$. The function $\lambda(\tau)$ satisfies the following conditions:

(L.1) $\lambda(\tau)$ is symmetric, with $\lambda(0) = 0$.

(L.2) $\lambda(\tau)$ is continuous and strictly convex for all τ in the interval $(-\tau_0, \tau_0)$, where $0 < \tau_0 \le \infty$. If $\tau_0 < \infty$, then $\lambda(\tau) = \infty$ for all τ outside the interval $(-\tau_0, \tau_0)$.

(L.3) $\lambda(\tau)$ is differentiable for all $\tau \in (-\tau_0, \tau_0)$ except possibly at $\tau = 0$.

(L.4) $\lambda(\tau) \to \infty$ and $\partial\lambda(\tau)/\partial\tau \to \infty$ as $\tau \to \tau_0$.

The symmetry condition in L.1 is not essential, but simplifies some of the proofs. To give a simple example of a function λ satisfying these conditions, let

$$\lambda_0(\tau;\sigma) = -\log[(1-\sigma|\tau|)^+] \tag{3.3}$$

where $x^+ \equiv \max\{x, 0\}$, and $\sigma > 0$ is a parameter. Then $\tau_0 = 1/\sigma$.

The functions H_k defined by (3.2) clearly satisfy conditions G.1, G.2, and G.4, but they do not satisfy condition G.3 for $k \ge 3$ because the maximum function does not have the needed higher-order derivatives. This means that there is not a density function corresponding to the function *F* in these cases. To show that *F* is in fact a distribution function, consider the GEV model defined by

$$H_k^*(y_1,...,y_k;t_1,...,t_k) = \int dt \left\{ \sum_{i=1}^{k} \left(y_i \exp[-\lambda(t_i - t)] \right)^\alpha \right\}^{1/\alpha} \tag{3.4}$$

with $\alpha > 1$. These functions H_k^* satisfy conditions G.1–G.4, so that the corresponding functions F^* defined by (3.1) are distribution functions. Now

$$\lim_{\alpha\to\infty}\left(\sum_i c_i^{\alpha}\right)^{1/\alpha} = \max_i \{c_i\} \tag{3.5}$$

for $c_i \geq 0$, $i = 1,...,k$, and convergence is uniform in $(c_1,...,c_k)$ provided the c_i are bounded. Therefore the integrand in (3.5) converges uniformly in t to the integrand of (3.2), and so $H_k^* \to H_k$ and $F^* \to F$ pointwise as $\alpha \to \infty$. The function F defined by (3.1) and (3.2) is then the limit of a sequence of distribution functions, and so is itself a (possibly defective) distribution function. But $H_k(0,...,0; t_1,...,t_k) = 0$ implies

$$F_{t_1,...,t_k}(\infty,...,\infty) = 1$$

and therefore F is a proper distribution function.

Now let W, the random part of the conditional indirect utility function (2.1), be a separable stochastic process on $[0,T]$ with finite-dimensional distribution functions defined by (3.1) and (3.2). Properties of W are given in the next four lemmas. Some results follow directly from the definition. The marginal distribution of $W(t)$ is type I extreme value,

$$F_t(w) = \exp(-c\, e^{-w})$$

where

$$c = 2\int_0^{\infty} d\tau\, e^{-\lambda(\tau)}, \tag{3.6}$$

which is finite by lemma A.1 (in the appendix). If $|t_1 - t_2| \geq 2\tau_0$, then $W(t_1)$ and $W(t_2)$ are independent. The correlation coefficient between $W(t_1)$ and $W(t_2)$ increases from 0 to 1 as $|t_1 - t_2|$ decreases from $2\tau_0$ to 0; its values can be computed numerically for any particular choice of λ.

Lemma 3.1. Let W be a stochastic process on $[0,T]$ with finite-dimensional distribution functions defined by (3.1) and (3.2), and suppose the function λ in (3.2) satisfies conditions L.1–L.4. Then W is continuous in probability.

Continuity in probability means that, for any given t, $\text{plim}[W(t+\delta) - W(t)] \to 0$ as $\delta \to 0$.

Lemma 3.2. Let W be a stochastic process satisfying the conditions of Lemma 3.1. Then W has the same finite dimensional distributions as the process W^* defined by

$$W^*(t) = \max_k \{X_k - \lambda(T_k - t)\} \tag{3.7}$$

where (X_k, T_k) is an enumeration of the points in the Poisson process on $\mathrm{R}\times[0,T]$ with intensity measure

$$(e^{-x}dx)\times ds\,. \tag{3.8}$$

This result is proved by de Haan (1984, theorem 3) for max-stable processes, a class of processes which includes the present model, under the assumption of continuity in probability.

The intensity measure (3.8) means that the number of points (or events) occurring in the time interval $(t, t+dt)$ and with levels in the interval $(x, x+dx)$ is a Poisson random variable with rate $e^{-x}dx\,dt$. The total number of events is infinite, because the total rate for the Poisson process is infinite, but the number of events with levels exceeding any given x is finite, with probability one.

De Haan calls (3.7) the spectral representation for max-stable processes. It has an obvious heuristic interpretation: the random component of utility is the upper envelope of a sequence of "bumps", with shape given by $-\lambda(t)$, occurring at random times and with random heights. However, a separable stochastic process is not uniquely determined by its finite-dimensional distributions, so W does not necessarily have the representation W^*.

Lemma 3.3. Let $W^*(t)$ and (X_k, T_k) be defined as in lemma 3.2. Then, with probability one, W^* has the representation

$$W^*(t) = \max_{k\in K}\{X_k - \lambda(T_k - t)\} > -\infty \tag{3.9}$$

where K is a (random) finite set of points of the Poisson process.

Lemma 3.4. Let W be a separable stochastic process satisfying the conditions of lemma 3.1. Then sample paths of W are continuous, with probability one.

The representation (3.9) actually implies stronger regularity properties for W^*: with probability one, sample paths are differentiable except at a (random) finite set of points.

Now we can derive the distribution of the maximum value of $U(t)$, using lemma 2.2. For any interval $J\subseteq[0,T]$, define the "envelope" function $q(\cdot\,;J)$ by

$$q(t;J)\equiv\sup_{s\in J}\{V(s)-\lambda(t-s)\}\,. \tag{3.10}$$

For the special case $J=[0,T]$, let

$$q(t) = q(t;[0,\,T])\,. \tag{3.11}$$

Lemma 3.5. Let $U(t) = V(t) + W(t)$, where V is a function satisfying conditions V.1 and V.2, and W is a separable stochastic process satisfying the conditions of lemma 3.1. Let $G(u\,;V)$ be the distribution function of maximum utility, defined by (2.8). Then

$$G(u\,;V) = \exp\left(-\,\mathrm{e}^{-u} \int dt\, e^{q(t)}\right) \tag{3.12}$$

with $q(t)$ defined by (3.10)–(3.11).

4. Choice probabilities: general formulation

A general expression for $P_0(J)$, where J is an open interval with respect to $[0,T]$, was found in Section 2 (lemma 2.3). Now we evaluate this expression when the stochastic process W is defined by the extreme-value model of Section 3.

Using the expression (3.12) for $G(u\,;V)$, the distribution function of maximum utility, we find

$$G(u\,;V - v\cdot 1\{J\}) = \exp\left(-\,e^{-u}\, K(J\,;v)\right) \tag{4.1}$$

where

$$K(J\,;v) = \int dt\, \exp\left(\max\left\{q(t\,;J) - v,\, q(t\,;J^c)\right\}\right). \tag{4.2}$$

J^c denotes the complement of J with respect to $[0,T]$, and $q(t\,;J)$ is defined by (3.10). Because the exponent in (4.1) factorizes into a term depending only on u and a term depending only on v, it is clear that the uniform differentiability condition of lemma 2.3 will be satisfied provided only that $K(J\,;v)$ is right-differentiable with respect to v at $v = 0$. Suppose that

$$K_1(J) = \frac{\partial}{\partial v^+}\left[K(J\,;v)\right]_{v=0} \tag{4.3}$$

does in fact exist (as will be shown). The plus sign indicates the right derivative. Then

$$\begin{aligned} G_1(u) &= \frac{\partial}{\partial v^+}\left[G(u\,;V - v\cdot 1\{J\})\right]_{v=0} \\ &= -\,G(u\,;V)\, e^{-u}\, K_1(J)\,. \end{aligned} \tag{4.4}$$

The dependence on u is now explicit, so the integration in (2.14) can be done directly. The result is

$$P_0(J) = -\frac{K_1(J)}{\int dt\, e^{q(t)}}\,, \tag{4.5}$$

where we have used the identity

$$K(J;0)=\int dt\, e^{q(t)}\,.$$

To investigate the differentiability of the integral $K(J;v)$, we define

$$R(J;v)=\{t \mid q(t;J)>q(t;J^c)+v \text{ and } q(t)>-\infty\} \tag{4.6}$$

for $v\ge 0$, and also $R(J)=R(J;0)$. Then R is an interval, as given by the next lemma.

Angle brackets will be used to denote intervals that may or may not include their end points. For example, $(a,b\rangle$ means that the interval is either (a,b) or $(a,b]$. We adopt the convention that (a,b), $(a,b]$ and $[a,b)$ are null if $a\ge b$, and $[a,b]$ is null if $a>b$.

Lemma 4.1. Suppose J is the open interval (a,b), with $0\le a<b\le T$, and $R(J;v)$ is defined as in (4.6). Assume that conditions V.1–V.2 and L.2–L.4 hold. Then there are functions $\eta(t,v)$ and $\zeta(t,v)$, right-continuous in v at $v=0$, such that

$$R\big((a,b);v\big)=\langle \eta(a,v),\, \zeta(b,v)\rangle \tag{4.7}$$

when $v\ge 0$.

If J is an open interval with respect to $[0,T]$, there are some special cases where J is not of the form (a,b), but the method used to prove lemma 4.1 can still be applied. For convenience of notation, we rewrite the intervals $[0,b)$, $(a,T]$, and $[0,T]$ as $(0-,b)$, $(a,T+)$, and $(0-,T+)$, where $0<b\le T$ and $0\le a<T$. Then (4.7) holds for these cases also, if we define

$$\eta(0-,v)=-\tau_0 \quad\text{and}\quad \zeta(T+,v)=T+\tau_0\,. \tag{4.8}$$

The remaining case is trivial: if $J=\varnothing$ then $R(J;v)=\varnothing$.

Using (4.7), we can rewrite (4.2) as

$$K(J;v)=\int_{-\tau_0}^{\eta(a,v)} dt\exp[q(t;J^c)]+e^{-v}\int_{\eta(a,v)}^{\zeta(b,v)} dt\exp[q(t;J)] \tag{4.9}$$

$$+\int_{\zeta(b,v)}^{T+\tau_0} dt\exp[q(t;J^c)]$$

in the case where $\zeta(b,v)\ge \eta(a,v)$, and

$$K(J;v)=\int dt\exp[q(t;J^c)] \tag{4.10}$$

otherwise.

Lemma 4.2. Suppose $J=(a,b)\subseteq[0,T]$, and $K(J;v)$ is defined by (4.2). Then, under the conditions of lemma 4.1, $K(J;v)$ is right-differentiable with respect to v at $v=0$, with derivative

$$K_1(J)=-\int_{\eta(a)}^{\zeta(b)} dt\, e^{q(t)} \tag{4.11}$$

where

$$\eta(a)=\eta(a,0) \quad \text{and} \quad \zeta(b)=\zeta(b,0)$$

in terms of the functions $\eta(t,v)$ and $\zeta(t,v)$ defined in lemma 4.1.

In the proof of lemma 4.2 we show that $\eta(a)\le\zeta(b)$, so the integral in (4.11) is properly defined. From the proof of lemma 4.1, we see that $\eta(a)$ is the "cross-over" point between $q(t;[0,a])$ and $q(t;(a,T])$, i.e., $t=\eta(a)$ is the solution of

$$q(t;[0,a])=q(t;(a,T]).$$

Similarly, $\zeta(b)$ is the cross-over point between $q(t;[0,b))$ and $q(t;[b,T])$. As before, the cases where J has the form $[0,b)$, $(a,T]$, or $[0,T]$ can be handled in the same way, using (4.8) for the appropriate limits of integration. If $J=\varnothing$ then $K_1=0$.

From (4.5) and (4.11), we finally get

$$P_0\big((a,b)\big)=\int_{\eta(a)}^{\zeta(b)} dt\, e^{q(t)} \bigg/ \int dt\, e^{q(t)}. \tag{4.12}$$

This is similar to the continuous logit model, but with the envelope function q replacing the systematic utility V and with $R((a,b))$ replacing the range of integration (a,b).

The next step is to find $P_0(J)$ for any interval $J\subset[0,T]$, and show that it is equal to the choice probability $P(J)$. We start by finding $P_0(J)$ for a closed interval, using

$$P_0([a,b])=\lim_{n\to\infty} P_0((a-1/n,b+1/n)). \tag{4.13}$$

Lemma 4.3. Suppose the stochastic process W is defined by the extreme-value model of section 3, and conditions V.1 and V.2 hold. Then

$$P_0([a,b])=\int_{\zeta(a)}^{\eta(b)} dt\, e^{q(t)} \bigg/ \int dt\, e^{q(t)} \tag{4.14}$$

for $0<a\le b<T$, where η and ζ are defined in lemma 4.2.

The method used to prove lemma 4.3 can also be used for intervals open at one end and closed at the other. If $J=(a,b]$ with $0\le a<b<T$ then the limits of integration are

$\eta(a)$ and $\eta(b)$, while if $J=[a,b)$ with $0<a<b\le T$ they are $\zeta(a)$ and $\zeta(b)$. To keep the notation consistent, we define

$$\zeta(0)=\eta(0-)=-\tau_0 \quad \text{and} \quad \eta(T)=\zeta(T+)=T+\tau_0 . \tag{4.15}$$

For any interval $J\subset[0,T]$, we have $J^c=J_1^c\cup J_2^c$, where J_1^c and J_2^c are disjoint intervals (or null). Using the representation of P_0 that we have just found, it is clear that

$$P_0(J)+P_0(J_1^c)+P_0(J_2^c)=1 . \tag{4.16}$$

But from the definitions (2.6) and (2.7), $P_0(J)\le P(J)$. This is compatible with (4.16) only if

$$P(J)=P_0(J) , \tag{4.17}$$

which also implies that a maximum utility "tie" between any two intervals has zero probability.

For all practical purposes, $P(J)$ needs to be defined only when the choice set J is an interval. However, it may be of some theoretical interest to show that $P(J)$ extends to a probability measure over the real interval $[0,T]$. This requires the following countable additivity property.

Lemma 4.4. Suppose $J_1,J_2,\ldots$ is a sequence of disjoint intervals of $[0,T]$ such that $\bigcup_{n=1}^{\infty} J_n=J$ is also an interval of $[0,T]$. Then, under the conditions of lemma 4.3,

$$P_0(J)=\sum_{n=1}^{\infty} P_0(J_n) . \tag{4.18}$$

From (4.18), and the results $P_0(\varnothing)=0$ and $P_0([0,T])=1$, it follows that P_0 defines a probability measure on the class of finite disjoint unions of intervals of $[0,T]$. Then by standard results (see, for example, Billingsley, 1986, section 3) P_0 has a unique extension to a probability measure on the class of Borel sets in $[0,T]$.

5. Choice probabilities: special cases

In this section we discuss three special cases that can arise in the choice probability model given by (4.12) and (4.14): (1) intervals that are never chosen; (2) discrete points that are chosen with nonzero probability; and (3) intervals on which a choice probability density can be defined.

(1) If $R((a,b))$ is empty or consists of a single point, or equivalently if $\eta(a)=\zeta(b)$, then $P((a,b))=0$. This typically occurs if at some nearby point s the parametric utility

$V(s)$ is much larger than its value in the interval (a,b). If some point $t \in (a,b)$ is to be chosen, then the path of the stochastic process *W* must rise steeply from *s* to *t*, so that $U(t) > U(s)$, but must not be too steep at *t*, so that $U(t)$ can be a maximum. According to lemma 3.3, the sample paths of *W* are made up of segments of the function λ, so the required curvature may not be attainable.

We can say that such an interval, which is never chosen, is in the "shadow" of some nearby peak of the function *V*. Examples are given in the next section.

(2) If $R([t,t])$ is a non-degenerate interval, or equivalently if $\eta(t) > \zeta(t)$, then the probability distribution of t^* has a discrete component $P(\{t\}) > 0$. This can happen only at points where *V* is irregular, as follows.

Lemma 5.1. Under the conditions of lemma 4.3, a necessary condition for $\eta(t) > \zeta(t)$ is that *V* is not differentiable at *t*. If *V* has (unequal) left and right derivatives at *t*, then further necessary conditions are that $V'(t+) < V'(t-)$ and $\max\{-V'(t+), V'(t-)\} \geq \lambda'(0+)$.

Although these conditions are necessary, they are not sufficient because *t* may be in the shadow of a higher value of *V* elsewhere. Typically there will be discrete choice probabilities at points where *V* has a step or a sharp peak, as in the examples in the next section.

Restricting the choice set to $[0,T]$ is equivalent to setting $V(t) = -\infty$ outside that interval. Consequently there are also discrete probabilities associated with the end-points $t = 0$ and $t = T$, except in the special case where $V(t) \to -\infty$ as *t* approaches the end-points. Heuristically, these discrete probabilities correspond to corner solutions of the utility maximization problem, where utility is decreasing at $t = 0$ or increasing at $t = T$. In some applications this might not be the appropriate way to model the end-point effects, in which case one would use only the conditional probabilities for an interior solution, $\Pr\{t^* \in J \mid t^* \in (0,T)\}$. The assumption about how *V* behaves at the end-points is obviously less important if the utility can be defined over a larger region than the interval on which (conditional) choice probabilities are needed.

(3) Under some additional assumptions, we can find a probability density function for t^*, i.e., $p(t) = dP([0,t])/dt$.

(D.1) The derivative $V'(t)$ exists at *t*.

(D.2) If $|V'(t)| \geq \lambda'(0+)$, then the second derivative $V''(t)$ exists at *t*.

(D.3) The second derivative $\lambda''(t)$ exists for all $t \in (0, \tau_0)$.

(D.4) There is a neighborhood N_t of t such that $P(J) > 0$ for all proper intervals $J \subset N_t$ (i.e., there is no shadowing in N_t).

Lemma 5.2. Suppose conditions D.1–D.4 hold at some point *t*. Then, under the conditions of lemma 4.3,

$$\frac{d\eta(t)}{dt} = 1 - \frac{V''(t)}{\lambda''(\eta(t)-t)} \qquad \text{if } |V'(t)| > \lambda'(0+), \tag{5.1}$$

$$\frac{d\eta(t)}{dt} = 1 \qquad \text{if } |V'(t)| < \lambda'(0+). \tag{5.2}$$

Differentiating $P([0,t])$, as given by (4.14), we find the probability density function

$$p(t) = \frac{d\eta}{dt} \exp[q(\eta)] \Big/ \int ds \exp[q(s)] \tag{5.3}$$

where the derivative of $\eta = \eta(t)$ is given by (5.1) or (5.2).

If $|V'(t)| = \lambda'(0+)$, then $p(t)$ may be discontinuous at *t*. From the proof of lemma 5.2, it is clear that we get (5.1) on the side where $|V'(t)|$ is larger, and (5.2) on the side where $|V'(t)|$ is smaller.

In the case $|V'(t)| < \lambda'(0+)$, we have $\eta(t) = t$. This implies $q(\eta) = V(t)$, so the probability density reduces to

$$p(t) = e^{V(t)} \Big/ \int ds\, e^{q(s)}. \tag{5.4}$$

If (5.4) holds for all *t* in some interval *J*, then the choice probability function conditional on *J* reduces to the continuous logit model,

$$p(t|J) = e^{V(t)} \Big/ \int_J ds\, e^{V(s)} \tag{5.5}$$

for $t \in J$. This shows that, under suitable restrictions, the continuous logit model is after all compatible with maximization of a stochastic process with (almost surely) continuous sample paths. The simplified model (5.5) is, however, unlikely to be useful for estimation because *V* will be a function of unknown parameters θ, and the condition that $|V'(t)| \le \lambda'(0+)$ for all $t \in J$ will hold only for restricted values of θ.

6. Construction and examples of choice probabilities

In practice the characteristics $x(t)$ are not observed as functions of *t* over the interval $[0,T]$. Instead, they are usually interpolated from observations at some finite set of points. We shall assume linear interpolation, for simplicity — quadratic or other forms of interpolation would make some of the following expressions algebraically more complicated, but the choice probabilities would still be computationally tractable. For the same reason we assume that the non-stochastic component of the utility function is linear in $x(t)$,

$$V(t) = V\big(x(t), z, \theta\big) = \sum_i \beta_i(z,\theta)\, x_i(t)\,. \tag{6.1}$$

Then $V(t)$ is piecewise linear in *t*. We should allow for discontinuities in $V(t)$: time-varying prices, for example, are typically step functions. Let $I_i = (t_i, t_{i+1})$, $i = 1,...,m$, (with $t_0 = 0$ and $t_{m+1} = T$) be a set of intervals on each of which $V(t)$ is linear,

$$V(t) = a_i + b_i(t - t_i)\,, \qquad t_i < t < t_{i+1}\,. \tag{6.2}$$

If V is discontinuous at $t = t_i$ then, according to condition V.2, $V(t_i)$ is the larger of $a_{i-1} + b_{i-1}(t_i - t_{i-1})$ and a_i. According to condition V.3, the coefficients a_i and b_i are continuous in θ.

The examples which follow are based on the particular choice (3.3) for the function λ in the GEV model of section 3:

$$\lambda(\tau) = -\log(1 - \sigma\tau) \qquad \text{for} \qquad 0 \le \tau < \tau_0 \tag{6.3}$$

with $\sigma > 0$ and $\tau_0 = 1/\sigma$. Computation of the choice probabilities hinges on the envelope function *q* defined by (3.11), so first we need a systematic procedure for constructing *q*. We have

$$q(s) = \max_i q(s; I_i) \tag{6.4}$$

where

$$q(s; I_i) = \max_{t_i \le t \le t_{i+1}} \left\{ a_i + b_i(t - t_i) - \lambda(s - t) \right\}. \tag{6.5}$$

Let $\psi_i(s)$ be the maximizing value of *t*, so that

$$q(s; I_i) = a_i + b_i\big(\psi_i(s) - t_i\big) - \lambda\big(s - \psi_i(s)\big). \tag{6.6}$$

When λ is given by (6.3), the solution is

$$\psi_i(s) = \begin{cases} t_i & \text{if} \quad s \le t_i - d_i \\ s + d_i & \text{if} \quad t_i - d_i \le s \le t_{i+1} - d_i \\ t_{i+1} & \text{if} \quad s \ge t_{i+1} - d_i \end{cases} \tag{6.7}$$

where

$$d_i = \operatorname{sgn}(b_i)\left(\sigma^{-1} - |b_i|^{-1}\right)^+ \tag{6.8}$$

(with $d_i = 0$ if $b_i = 0$). From (6.6) and (6.7), we see that $q(s;I_i)$ is linear for $s \in [t_i - d_i, t_{i+1} - d_i]$ and is log linear outside that interval, with a maximum at t_{i+1} if $b_i > 0$ or at t_i if $b_i < 0$. For example, if $b_i \geq \sigma$ then

$$q(s;I_i) = \begin{cases} a_i + b_i(t_{i+1} - t_i) + \log[1 - \sigma|s - t_{i+1}|] & \text{if } t_{i+1} - d_i \leq s < t_{i+1} + \sigma^{-1} \\ a_i + b_i(s - t_i + d_i) - \log(b_i/\sigma) & \text{if } t_i - d_i \leq s \leq t_{i+1} - d_i \\ a_i + \log[1 - \sigma(t_i - s)] & \text{if } t_i - \sigma^{-1} < s \leq t_i - d_i \ . \end{cases} \tag{6.9}$$

For each i and $j \neq i$, define

$$M_{i,j} = \{s \mid q(s;I_i) \geq q(s;I_j) \text{ and } q(s;I_i) > -\infty\}. \tag{6.10}$$

By the argument used to prove lemma 4.1, $M_{i,j}$ is an interval. In fact it is always a non-degenerate interval because there is an interval where $q(s;I_i) > -\infty$ but $q(s;I_j) = -\infty$. Then

$$M_{i,j} = (t_i - \sigma^{-1}, m_{i,j}\rangle \qquad \text{if } i < j$$

$$M_{i,j} = \langle m_{i,j}, t_{i+1} - \sigma^{-1}) \qquad \text{if } i > j\,.$$

The point $m_{i,j}$ is found by comparing each of the segments of $q(s;I_i)$ (as in equation 6.9) with each of the segments of $q(s;I_j)$ to determine if and where they intersect or, possibly, coincide. Then we can determine the (possibly degenerate) intervals

$$S_i = \{s \mid q(s;I_i) = q(s)\} = \bigcap_{j \neq i} M_{i,j}\ . \tag{6.11}$$

The envelope function can be computed as

$$q(s) = \sum_{i=1}^{m} q(s;I_i)\, 1[s \in \{S_i - S_{i+1}\}] \tag{6.12}$$

(with $S_{m+1} = \varnothing$), which allows for the possibility that adjacent intervals may overlap.

Having found q, we can determine the discrete probability points, if any. The only candidates are the points t_i, $i = 0,1,\ldots,m+1$. If V is continuous at t_i, then

$$R(\{t_i\}) = \operatorname{int}\{S_{i-1} \cap S_i\}\,. \tag{6.13}$$

If V is discontinuous at t_i, then

$$R(\{t_i\}) = \begin{cases} (t_i - \sigma^{-1}, t_i - d_i) \cap \operatorname{int} S_i & \text{if } V(t_i) = V(t_i +) \\ (t_i - d_{i-1}, t_i + \sigma^{-1}) \cap \operatorname{int} S_{i-1} & \text{if } V(t_i) = V(t_i -) . \end{cases} \tag{6.14}$$

If $R(\{t_i\})$ is non-degenerate, i.e., $\zeta(t_i) < \eta(t_i)$, then t_i has a discrete choice probability $P([t_i, t_i])$ given by (4.14).

Elsewhere in $(0,T)$ there is a probability density given by (5.3). If $t_i \le t \le t_{i+1}$, and if t is not a discrete probability point, then

$$p(t) = \begin{cases} \dfrac{\exp[q(t-d_i)]}{\int ds \exp[q(s)]} & \text{if } t - d_i \in S_i \\ 0 & \text{otherwise} . \end{cases} \tag{6.15}$$

Example 1 ("step")

$$V(t) = \begin{cases} 0 & \text{if } 0 \le t < T/2 \\ c & \text{if } T/2 \le t \le T . \end{cases} \tag{6.16}$$

For this example, consider the case where $c \ge 0$ and $\sigma^{-1} \le T/2$. Then $R(\{0\}) = (-\sigma^{-1}, 0)$, $R(\{T/2\}) = (t_1, T/2)$, and $R(\{T\}) = (T, T + \sigma^{-1})$, corresponding to discrete probabilities, with

$$t_1 = T/2 - \sigma^{-1}(1 - e^{-c}) .$$

The resulting probabilities are

$$P(\{0\}) = 1/(2\sigma D_1),$$

$$P(\{T/2\}) = (e^c - e^{-c})/(2\sigma D_1),$$

$$P(\{T\}) = e^c/(2\sigma D_1),$$

and

$$p(t) = \begin{cases} 1/D_1 & \text{if } 0 < t \le t_1 \\ 0 & \text{if } t_1 < t < T/2 \\ e^c/D_1 & \text{if } T/2 < t < T \end{cases}$$

where

$$D_1 = (T/2)(1 + e^c) + (2e^c + e^{-c} - 1)/(2\sigma) .$$

The interval $(t_1, T/2)$ is in the "shadow" of the step at $t = T/2$.

Example 2 ("peak")

$$V(t) = \begin{cases} 2ct/T & \text{if } 0 \le t \le T/2 \\ 2c(T-t)/T & \text{if } T/2 \le t \le T \end{cases} \tag{6.17}$$

where $c > 0$. For this example, consider the case where $2c/T > \sigma$, so that there is nonzero discrete probability at the peak. Then

$$P(\{0\}) = P(\{T\}) = \sigma T/(8c^2 D_2),$$

$$P(\{T/2\}) = \frac{e^c}{D_2}\left(\frac{1}{\sigma} - \frac{\sigma T^2}{4c^2}\right),$$

and

$$p(t) = \begin{cases} \dfrac{\sigma T}{2cD_2} \exp\left(\dfrac{2ct}{T}\right) & \text{if } 0 < t < T/2 \\ \dfrac{\sigma T}{2cD_2} \exp\left(\dfrac{2c(T-t)}{T}\right) & \text{if } T/2 < t < T \end{cases}$$

where

$$D_2 = \frac{e^c}{\sigma} + \frac{\sigma T^2}{4c^2}(e^c - 1).$$

Apart from a constant factor, this density function is the same as in the continuous logit model, but the discrete probability components are of course different. If the peak at $t = T/2$ were "rounded", so that $V'(t)$ were continuous there, then the discrete component at $t = T/2$ would disappear. For a rounded peak, however, the Jacobian (5.1) would no longer be constant, so the density $p(t)$ would differ from that of the continuous logit model.

Example 3 ("valley")

$$V(t) = \begin{cases} -2ct/T & \text{if } 0 \le t \le T/2 \\ -2c(T-t)/T & \text{if } T/2 \le t \le T \end{cases} \tag{6.18}$$

where $c > 0$. For this example consider the case where $2c/T > \sigma > 2/T$ (which implies $c > 1$). In this case there is shadowing of some, but not all, of the "valley". The resulting probabilities are

$$P(\{0\}) = P(\{T\}) = \frac{1}{D_3}\left(\frac{1}{\sigma} - \frac{\sigma T^2}{8c^2}\right)$$

and

$$p(t) = \begin{cases} \frac{\sigma T}{2cD_3}\exp\left(-\frac{2ct}{T}\right) & \text{if } 0 < t < \frac{T}{2} - d_3 \\ 0 & \text{if } \frac{T}{2} - d_3 < t < \frac{T}{2} + d_3 \\ \frac{\sigma T}{2cD_3}\exp\left(-\frac{2c(T-t)}{T}\right) & \text{if } \frac{T}{2} + d_3 < t < T \end{cases}$$

where

$$d_3 = \sigma^{-1} - \frac{T}{2c}$$

$$D_3 = \frac{2}{\sigma} + \frac{\sigma T^2}{4c^2} + \frac{\sigma T^2}{2c^2}\exp\left(-c + \frac{2c}{\sigma T} - 1\right).$$

In this example the shadowing is due to the steepness of the sides of the valley, and would generally happen even if the piecewise linear V were replaced by a smoother function.

7. Parameter estimation: discrete observations

Going back to the general case, we briefly consider the problem of estimating the parameter vector θ in (6.1) from a random sample of observations on x, z and t. In practice the choice t^* is not usually observed as a continuous variable. Instead, the choice set $[0,T]$ is partitioned into a finite set of intervals J_i, $i = 1,...,n$, and we observe the index i of the chosen interval (i.e., $t^* \in J_i$). The intervals J_i need not be the same as the intervals I_i used in the interpolation of $V(t)$; in fact there is some simplification if they are not the same because then any discrete probability point (other than 0 and T) is in the interior of a choice interval J_i.

In that case the problem reduces to estimation of a discrete choice model, with

$$P(i \mid x, z, \theta) = P(J_i)\,. \qquad (7.1)$$

Although we are back to the discrete case, it is a discrete choice model in which the stochastic utility has a covariance structure compatible with an underlying continuous-time stochastic process that has almost surely continuous sample paths. Estimation from data where t^* itself is observed will be discussed elsewhere.

A discrete choice model like (7.1) is usually estimated by maximum likelihood estimation. Let Θ denote the parameter space, N the sample size, and $L_{m,N}(\theta)$ the log likelihood function for m discrete alternatives. Then $\hat{\theta}_N$ is determined (not necessarily uniquely) by

$$L_{m,N}(\hat{\theta}_N) = \max_{\theta \in \Theta} L_{m,N}(\theta)$$

if a maximum exists or, if not, $\hat{\theta}_N$ is a limit point of a sequence $\{\theta_{N,i}\}$ satisfying

$$\lim_{i \to \infty} L_{m,N}(\theta_{N,i}) = \sup_{\theta \in \Theta} L_{m,N}(\theta).$$

Such an estimator exists almost surely, provided $\hat{\theta}_N$ is allowed to be on the boundary of Θ, including if necessary a point at infinity. Sufficient sets of conditions for strong consistency of $\hat{\theta}_N$ are given by several authors. In particular, see conditions (1)–(4) in McFadden (1984, section 3.2). These conditions do not require the probabilities to be bounded away from zero.

Although the consistency conditions are quite mild, the requirement that $P(i \mid x, z, \theta)$ should be continuous in θ (except possibly on a set of (x, z) with zero probability) has to be checked in any particular model. From assumption V.4, and from the definition of q as an envelope function, it can readily be shown that $q(s)$ is continuous in θ uniformly in s, and that $\int ds \exp[q(s)]$ is continuous in θ. However, we also need to show that the limits of integration η and ζ are continuous in θ. The following additional assumptions will be used.

(C.1) The second derivative $\lambda''(\tau)$ exists for all $\tau \in (0, \tau_0)$ and has a strictly positive lower bound.

(C.2) The function $\Delta(t_1, t_2, \theta)$ defined by

$$V(t_1, \theta) - V(t_2, \theta) = (t_1 - t_2)\, \Delta(t_1, t_2, \theta) \qquad (7.2)$$

is continuous in θ uniformly in t_1 and t_2 for all $t_1 \in [t_0, T]$ and all $t_2 \in [0, t_0]$, where t_0 is some specified point in $[0, T]$.

Lemma 7.1. Suppose that conditions V.1–V.4, L.1–L.4, and C.1–C.2 hold. Then $\eta(t_0)$ and $\zeta(t_0)$ are continuous in θ.

The uniform continuity condition (7.2) will be satisfied by most parametric models for $V(t,\theta)$, provided of course that $t = t_0$ is not one of the points where V can be discontinuous in t. The model discussed in section 6, where V is given by (6.2) and λ by (6.3), clearly satisfies the conditions of lemma 7.1 provided that the steps in V (if any) are not at the boundaries of choice intervals J_i. In fact, this proviso is not necessary for the model of section 6: even if $V(t,\theta)$ is discontinuous in t at $t = t_0$ for some values of θ, the cross-over points $\eta(t_0)$ and $\zeta(t_0)$ can be determined (by enumerating all the possible cases) and are found to be continuous in θ. In that model, therefore, the $P(i \mid x, z, \theta)$ are always continuous in θ.

In principle it is also important to check the identification condition (condition (4) in McFadden, 1984, Section 3.2), but it is generally quite difficult to find verifiable conditions that make θ identifiable. Even in the model defined by (6.2) and (6.3), one can construct pathological examples in which, say, $P(i \mid x, z, \theta) = 1$ for a range of θ. The usual approach is to suppose a discrete model to be identified except when there is an obvious transformation of θ that leaves the expected log likelihood invariant.

The choice probabilities $P(i \mid x, z, \theta)$ are generally not differentiable with respect to θ in a neighborhood of θ_0, even after excluding a set of (x, z) with probability zero. In Example 2 of section 6, for example, $P([0, T/2])$ is continuous but not differentiable with respect to $c = c(x, z, \theta)$ at $c = \sigma T/2$. As θ varies over a neighborhood of θ_0, the values of (x, z) at which $c = \sigma T/2$ will in general range over a set with nonzero probability. This lack of differentiability means that the usual theorems on asymptotic normality of $\hat{\theta}_N$ are not applicable, and further investigation of the asymptotic properties of $\hat{\theta}_N$ is needed. It may also be necessary to use gradient-free methods to maximize $L_N(\theta)$.

Appendix

The appendix contains proofs of the lemmas in the text (except where a specific reference is given), together with some preliminary results that are needed in the proofs.

Maximum of a stochastic utility function

Proof of Lemma 2.1. First, suppose that J is an open interval with respect to $[0,T]$ (i.e., J is an open interval with respect to R, except that if an end-point is 0 or T then that end-point may be included). The complement of J with respect to $[0,T]$, denoted J^c, is the union of at most two closed intervals. Because $U(t)$ is upper semicontinuous and J^c is compact,

$$\sup_{s\in J^c} U(s) = U(s_1) \text{ for some } s_1 \in J^c,$$

and therefore

$$\begin{aligned}\{M \subseteq J\} &= \{U(t) > \sup_{s\in J^c} U(s) \text{ for some } t\in J\} \\ &= \bigcup_{u\in \mathrm{R}_0} \left(\{U(t) > u \text{ for some } t\in J\} \cap \{\sup_{s\in J^c} U(s) \le u\} \right)\end{aligned} \tag{A.1}$$

where R_0 is the set of rational numbers. Using the separability of U with respect to D, we can rewrite this as

$$\{M \subseteq J\} = \bigcup_{u\in \mathrm{R}_0} \left(\left[\bigcup_{t\in J\cap D} \{U(t) > u\} \right] \bigcap \left[\bigcup_{s\in J^c \cap D} \{U(s) \le u\} \right] \right), \tag{A.2}$$

which is an element of $\mathscr{R}^{[0,T]}$.

Next suppose that J is a closed interval, say $[a,b]$ with $a \le b$. We have

$$\{M \subseteq J\} = \bigcap_{n=1}^{\infty} \{M \subseteq J_n\} \tag{A.3}$$

where

$$J_n = (a - 1/n, b + 1/n) \cap [0,T]. \tag{A.4}$$

Each of the sets $\{M \subseteq J_n\}$ is of the form (A.2), and therefore (A.3) is also an element of $\mathscr{R}^{[0,T]}$. A similar argument completes the proof for the case where the interval J is open at one end and closed at the other. □

Proof of Lemma 2.2. Let W be separable with respect to some countable dense set D. Without loss we can choose D to contain the discontinuity points of V, so U is also separable with respect to D. Let J_n be a sequence of finite sets such that $J_n \uparrow J \cap D$. Then

$$\left\{ \sup_{t \in J_n} U(t) \leq u \right\} \downarrow \left\{ \sup_{t \in J \cap D} U(t) \leq u \right\}$$

and therefore

$$\Pr\left\{ \sup_{t \in J_n} U(t) \leq u \right\} \downarrow \Pr\left\{ \sup_{t \in J \cap D} U(t) \leq u \right\} = \Pr\left\{ \sup_{t \in J} U(t) \leq u \right\}, \tag{A.5}$$

where the last equality follows from separability. Rewriting the left-hand side of (A.5) as

$$\Pr\{W(t) \leq u - V(t) \text{ for all } t \in J_n\} = F_{\{J_n\}}[u - V(t_1), \ldots, u - V(t_{m(n)})]$$

then gives the result (2.9). □

Proof of Lemma 2.3. For any $B > 0$, divide $(-B, B)$ into intervals of width $\varepsilon_N = 2^{-N}$ $(N = 1, 2, \ldots)$ and set $v_{i,N} = i\varepsilon_N$ $(i = 0, \pm 1, \ldots, \pm m_N)$ where $m_N = B/\varepsilon_N$. Define the event

$$A_{i,N} = \{U_1 \in (v_{i,N}, v_{i+1,N}] \text{ and } U_2 < v_{i,N}\} \tag{A.6}$$

where

$$U_1 = \sup_{t \in J} U(t), \qquad U_2 = \sup_{t \in J^c} U(t).$$

Then

$$\Pr\{U_1 > U_2 \text{ and } -B < U_1 \leq B\} = \lim_{N \to \infty} \sum_{i=-m_N}^{m_N - 1} \Pr\{A_{i,N}\}. \tag{A.7}$$

As in (2.11), we have

$$\Pr\{A_{i,N}\} = G(v_{i,N}; V - \varepsilon_N 1\{J\}) - G(v_{i,N}; V). \tag{A.8}$$

Given a point $u \in (-B, B]$, define $j = j(u, N)$ so that u is in the interval $(v_{j,N}, v_{j+1,N}]$. Applying the uniform differentiability condition (2.13) to the right hand side of (A.8),

$$\lim_{N \to \infty} \frac{1}{\varepsilon_N} \Pr\{A_{j(u,N),N}\} = G_1(u) \tag{A.9}$$

at each point u, and therefore, by the dominated convergence theorem,

$$\lim_{N\to\infty} \sum_{i=-m_N}^{m_N-1} \Pr\{A_{i,N}\} = \int_{-B}^{B} du\, G_1(u). \tag{A.10}$$

Letting $B \to \infty$ in (A.7) and (A.10) gives

$$\Pr\{U_1 > U_2\} = \int_{-\infty}^{\infty} du\, G_1(u). \tag{A.11}$$

According to the definition (2.6) this is just equal to $P_0(J)$, provided that *J* is an open interval. □

Properties of the continuous-time GEV model

The following preliminary lemmas establish useful results involving the function λ, which satisfies assumptions L.1–L.4.

Lemma A.1. The integral $\int d\tau \exp[-\lambda(\tau)]$ is finite.

Proof. Choose any point $\tau_1 \in (0, \tau_0)$. Then $0 < \lambda'(\tau_1) < \infty$. Strict convexity of λ implies

$$\lambda(\tau) > \lambda(\tau_1) + (\tau - \tau_1)\lambda'(\tau_1)$$

and therefore

$$\int_0^{\tau_0} d\tau\, e^{-\lambda(\tau)} < [\lambda'(\tau_1)]^{-1} \exp[-\lambda(\tau_1) + \tau_1 \lambda'(\tau_1)]$$

which is finite as required. □

Lemma A.2. If $y_1 > 0$, $y_2 > 0$, and $0 < |t_1 - t_2| < 2\tau_0$, then the equation

$$y_1 \exp[-\lambda(t - t_1)] = y_2 \exp[-\lambda(t - t_2)] \tag{A.12}$$

has exactly one nontrivial solution for *t*. (Trivial solutions are those with $\lambda(t - t_1) = \lambda(t - t_2) = \infty$.)

Proof. Suppose without loss that $t_1 < t_2$. Set $s = t_2 - t_1$, and $\tau = t - t_1$. If τ is outside the interval $(s - \tau_0, \tau_0)$ then one or both of $\lambda(\tau)$ and $\lambda(\tau - s)$ are infinite, so any solution would be a trivial one. If $\tau \in (s - \tau_0, \tau_0)$, we are looking for solutions of

$$f_s(\tau) = \log(y_1 / y_2) \tag{A.13}$$

where we have defined

$$f_s(\tau) = \lambda(\tau) - \lambda(\tau - s).$$

If τ_0 is finite, then $\lambda(\tau_0 - s)$ is finite, and so $f_s(\tau) \to \infty$ as $\tau \to \tau_0$ from below. If $\tau_0 = \infty$, then $\lambda'(\tau - s) \to \infty$ as $\tau \to \infty$ (from condition L.4); the inequality $f_s(\tau) > s\,\lambda'(\tau - s)$ (from strict convexity of λ) then shows that $f_s(\tau) \to \infty$ as $\tau \to \tau_0$ in this case also. By a similar argument, $f_s(\tau) \to -\infty$ as $\tau \to s - \tau_0$ from above. Thus the continuous function $f_s(\tau)$ is unbounded on the interval $(s - \tau_0, \tau_0)$, and therefore (A.13) has at least one solution in that interval.

If $s < \tau_0$ and τ is in one of the intervals $(s - \tau_0, 0)$ or (s, τ_0), then $f_s'(\tau) > 0$ because $\lambda'(\tau)$ is strictly increasing (from conditions L.2 and L.3). If $\tau \in (0, s)$ then conditions L.1–L.3 imply $\lambda'(\tau) > 0$ and $\lambda'(\tau - s) < 0$, so again $f_s'(\tau) > 0$. The function $f_s(\tau)$ is continuous on $(s - \tau_0, \tau_0)$, so the positive derivative implies that it is strictly increasing. Therefore (A.13) has at most one solution in that interval.

Combining these results, it follows that there is just one nontrivial solution. □

Lemma A.3. Let $\tau = \rho(\varepsilon, \delta)$ be the solution of

$$\lambda(\tau) + \varepsilon = \lambda(\tau + \delta), \qquad |\tau| < \tau_0 \tag{A.14}$$

for any given $\varepsilon > 0$ and $\delta \in (0, \tau_0)$. (By Lemma A.2, this solution exists and is unique.) Then

$$\int_{\rho(\varepsilon,\delta)}^{\infty} d\tau\, e^{-\lambda(\tau)} = o(\delta) \tag{A.15}$$

as $\delta \to 0+$.

Proof. First, note that the upper limit of integration can be changed to τ_0. Since $\lambda'(\tau)$ is strictly increasing (from conditions L.2 and L.3),

$$\lambda'(\rho) < (\varepsilon / \delta) < \lambda'(\rho + \delta),$$

where ρ is the solution of (A.14), and therefore $\lambda'(\rho + \delta) \to \infty$ as $\delta \to 0+$. Because $\lambda'(\tau) \to \infty$ as $\tau \to \tau_0$ (condition L.4) and is finite for $|\tau| < \tau_0$, it follows that $\rho + \delta \to \tau_0$ as $\delta \to 0+$.

From strict convexity of λ,

$$\begin{aligned}\lambda(\tau) &> \lambda(\rho + \delta) + (\tau - \rho - \delta)\,\lambda'(\rho + \delta) \\ &> \lambda(\rho + \delta) + (\tau - \rho - \delta)\,\varepsilon / \delta\end{aligned}$$

and therefore

$$\int_{\rho}^{\tau_0} d\tau\, e^{-\lambda(\tau)} < \exp[-\lambda(\rho+\delta)+\varepsilon] \int_{\rho}^{\tau_0} d\tau \exp[-(\tau-\rho)\varepsilon/\delta]$$
$$< \exp[-\lambda(\rho+\delta)+\varepsilon]\,(\delta/\varepsilon). \tag{A.16}$$

As shown above, $\rho+\delta \to \tau_0$ as $\delta \to 0+$, and therefore $\exp[-\lambda(\rho+\delta)] \to 0$. Therefore the left-hand integral in (A.16) is $o(\delta)$, as required. □

Remark. In lemma A.3 we had $\varepsilon > 0$ and $\delta > 0$. By symmetry,

$$\rho(\varepsilon,\delta) = -\rho(\varepsilon,-\delta) = -\rho(-\varepsilon,\delta) - \delta = \rho(-\varepsilon,-\delta) - \delta,$$

and so (A.15) holds also when $\varepsilon < 0$ and $\delta \to 0-$. If $\varepsilon > 0$ and $\delta \to 0-$, or if $\varepsilon < 0$ and $\delta \to 0+$, the corresponding result is

$$\int_{-\infty}^{\rho(\varepsilon,\delta)} d\tau\, e^{-\lambda(\tau)} = o(\delta). \tag{A.17}$$

Proof of Lemma 3.1. Fix *t*, and let $\delta_n \to 0$. First, consider the subsequence of strictly positive terms, $\delta_n > 0$. Define

$$F(w_1,w_2;\delta) = F_{t,t+\delta}(w_1,w_2)$$
$$= \exp\left(-\int d\tau \max\{\exp[-\lambda(\tau+\delta)-w_1], \exp[-\lambda(\tau)-w_2]\}\right). \tag{A.18}$$

As before, let $\tau = \rho(\varepsilon,\delta)$ be the solution of (A.14). When $\tau < \rho(w_2 - w_1,\delta)$ the first exponent in the integrand of (A.18) is larger, and when $\tau > \rho(w_2 - w_1,\delta)$ the second exponent is larger, so (A.18) becomes

$$F(w_1,w_2;\delta) = \exp\Bigg(-e^{-w_1}\int_{-\infty}^{\rho(w_2-w_1,\delta)} d\tau \exp[-\lambda(\tau+\delta)]$$
$$-e^{-w_2}\int_{\rho(w_2-w_1,\delta)}^{\infty} d\tau \exp[-\lambda(\tau)]\Bigg). \tag{A.19}$$

Now

$$\Pr\{|W(t)-W(t+\delta)|<\varepsilon\} = \int_{-\infty}^{\infty} du\,[F_1(u,u+\varepsilon;\delta) - F_1(u,u-\varepsilon;\delta)] \tag{A.20}$$

where

$$F_1(w_1,w_2;\delta) = \frac{\partial F(w_1,w_2;\delta)}{\partial w_1}. \tag{A.21}$$

Evaluating the integral in (A.20) gives

$\Pr\{|W(t)-W(t+\delta)|<\varepsilon\}=$

$$\frac{\int_{-\infty}^{\rho(\varepsilon,\delta)} d\tau \exp[-\lambda(\tau+\delta)]}{\int_{-\infty}^{\rho(\varepsilon,\delta)} d\tau \exp[-\lambda(\tau+\delta)]+\mathrm{e}^{-\varepsilon}\int_{\rho(\varepsilon,\delta)}^{\infty} d\tau \exp[-\lambda(\tau)]} \tag{A.22}$$

$$-\frac{\int_{-\infty}^{\rho(-\varepsilon,\delta)} d\tau \exp[-\lambda(\tau+\delta)]}{\int_{-\infty}^{\rho(-\varepsilon,\delta)} d\tau \exp[-\lambda(\tau+\delta)]+\mathrm{e}^{\varepsilon}\int_{\rho(-\varepsilon,\delta)}^{\infty} d\tau \exp[-\lambda(\tau)]}\ .$$

Using lemmas A.1 and A.3, we see that $\Pr\{|W(t)-W(t+\delta)|<\varepsilon\}\to 1$ as $\delta\to 0+$. The same result is obtained when $\delta\to 0-$, and the case $\delta=0$ is trivial. Thus for any sequence $\delta_n\to 0$,

$$\operatorname{plim}[W(t+\delta_n)-W(t)]=0$$

as required. □

Proof of Lemma 3.3. Suppose first that t is in the interval $J=[a,a+h]\subseteq[0,T]$, where $0<h<\tau_0$. Let $N_1(x)$ be the number of points of the Poisson process with $X_k>x$ and $T_k\in J$, and let $N_2(x)$ be the number of points with $X_k>x-\lambda(h)$. Then the random variables N_1 and N_2 have Poisson distributions with rates he^{-x} and $T\exp[\lambda(h)-x]$ respectively. Define the sequence of events

$$A(x_n,m_n)=\{0<N_1(x_n)\le m_n\}\cup\{N_2(x_n)\le m_n\} \tag{A.23}$$

for $n=1,2,...$ Choose x_n sufficiently small (i.e., large negative) such that $\Pr\{N_1=0\}<(3n)^{-1}$, and then m_n sufficiently large that $\Pr\{N_1>m_n\}$ and $\Pr\{N_2>m_n\}$ are both less than $(3n)^{-1}$, so that

$$\Pr\{A(x_n,m_n)\}>1-n^{-1}.$$

Then

$$\Pr\Big\{\bigcup_n A(x_n,m_n)\Big\}=1$$

so that, with probability one, at least one of the events $A(x_n,m_n)$ must occur.

Thus for some x_n there is at least one event with $X_k>x_n$ and $T_k\in J$. First, this implies $W^*(t)$ finite for $t\in J$, because the width of the interval is less than τ_0. Secondly it implies that, when $t\in J$, only events with $X_k>x_n-\lambda(h)$ can contribute to the maximum in (3.7). But there are a finite number (at most m_n) of such events. It follows that, with probability one, (3.9) holds for $t \in J$. Since $[0,T]$ can be covered with a finite number of intervals like *J*, (3.9) holds with probability one for all $t \in [0,T]$. □

Proof of Lemma 3.4. Consider the stochastic process W^* defined in lemma 3.2. The processes *W* and W^* have the same finite-sample distributions. According to lemma 3.3, almost every sample path $w^*(t)$ is the maximum of a finite number of continuous functions, i.e., $\{x_k - \lambda(t_k - t)\}$ with $k \in K$ and with $|t_k - t| < \tau_0$, and therefore $w^*(t)$ is continuous with probability one. Now if a stochastic process has continuous sample paths with probability one, then so do all other separable stochastic process with the same finite-dimensional distributions (see, for example, Billingsley, 1986, theorem 38.2), in particular *W*. □

The next results involve the "envelope functions" $q(t;J)$ and $q(t)$, defined by (3.10) and (3.11).

Lemma A.4. Suppose conditions V.1–V.2 and L.2 hold. Let $J \subseteq [0,T]$ be a closed interval. If $q(t;J) > -\infty$, then there is a point $s \in J$ such that $q(t) = V(s) - \lambda(t-s)$.

Proof. For some given *t* with $q(t;J) > -\infty$, let $s_1, s_2, \ldots$ be a sequence of points in *J* such that $V(s_i) - \lambda(t - s_i) \to q(t;J)$. Then we can choose a subsequence such that $s_i \to s \in J$. From conditions V.1–V.2 we have $V(s_i) \to V(s\pm) \le V(s) < \infty$, and from condition L.2 we have $\lambda(t - s_i) \to \lambda(t-s)$, so that $q(t;J) \le V(s) - \lambda(t-s)$. But from the definition (3.10) we must have $q(t;J) \ge V(s) - \lambda(t-s)$, and therefore $q(t;J) = V(s) - \lambda(t-s)$ as required. □

Proof of Lemma 3.5. In lemma 2.2, let *J* be the interval $[0,T]$ and let the distribution function *F* be given by (3.1) and (3.2). Then

$$G(u;V) = \lim_{n\to\infty} \exp\left(-\mathrm{e}^{-u} \int dt \exp[q(t;J_n)]\right) \tag{A.24}$$

where the function $q(t;J)$ is defined by (3.10), and the finite sets J_n $(n = 1,2,\ldots)$ satisfy $J_n \uparrow [0,T] \cap D$ where *D* is some countable dense set. Now consider

$$\lim_{n\to\infty} \int dt \exp[q(t;J_n)]. \tag{A.25}$$

Suppose $q(t) > -\infty$. From the definition (3.10), $A \subset B$ implies $q(t;A) \le q(t;B)$, so the integrand is monotonically increasing at each *t*. By lemma A.4, there is some $s = s(t) \in [0,T]$ such that

$$q(t) = V(s) - \lambda(t-s).$$

Given $\varepsilon > 0$, we can then find some $\delta > 0$ such that

$$[V(s) - \lambda(t-s)] - [V(s') - \lambda(t-s')] < \varepsilon$$

for all s' in at least one of the intervals $[s-\delta, s]$ and $[s, s+\delta]$, according to conditions V.2 and L.2. For n sufficiently large, there will be points of J_n in these intervals, and therefore

$$q(t) - q(t; J_n) < \varepsilon .$$

Thus $q(t; J_n) \uparrow q(t)$ for each t at which $q(t) > -\infty$. If $q(t) = -\infty$, then so is $q(t; J_n)$ for all n. The limit of (A.25) is then just $\int dt \exp[q(t)]$, by monotone convergence. Substituting this in (A.24), we get the result (3.12). □

Choice probabilities

Next, we define

$$\psi(s) = \{ t \mid V(t) - \lambda(s-t) = q(s),\ t \in [0,T] \}, \tag{A.26}$$

where s is such that $q(s) > -\infty$, and q is the envelope function (3.11). According to lemma A.4, $\psi(s)$ is non-empty. We shall need the set $R(J; v)$ defined by (4.6), and also the complementary set

$$R^*(J; v) = \left\{ t \mid q(t; J) \le q(t; J^c) + v \text{ and } q(t) > -\infty \right\}, \tag{A.27}$$

where $J \subseteq [0,T]$ is an interval. In the special case $v = 0$ we write $R(J) = R(J; 0)$ and $R^*(J) = R^*(J; 0)$.

Lemma A.5. Suppose $t_i \in \psi(s_i)$, where $q(s_i) > -\infty$ $(i = 1,2)$, and assume that the function λ satisfies conditions L.2–L.4. Then $t_1 > t_2$ implies $s_1 \ge s_2$.

Proof. First, define

$$S(t_1, t_2) = \{ s \mid V(t_1) - \lambda(s - t_1) \ge V(t_2) - \lambda(s - t_2) \text{ and } \lambda(s - t_1) < \infty \}. \tag{A.28}$$

If $0 < |t_1 - t_2| < 2\tau_0$, then from lemma A.2 there is just one solution of

$$V(t_1) - \lambda(s - t_1) = V(t_2) - \lambda(s - t_2)$$

for which $\lambda(s - t_1) < \infty$. Call the solution s_0. It is clear that $S(t_1, t_2)$ lies to the right of s_0 if $t_1 > t_2$, and to the left if $t_2 > t_1$. If $|t_1 - t_2| \ge 2\tau_0$, then $S(t_1, t_2)$ is just the region where $V(t_1) - \lambda(s - t_1)$ is finite. Thus $S(t_1, t_2)$ is always an interval:

$$S(t_1,t_2)=\begin{cases}[s_0,t_1+\tau_0) & \text{if } t_2<t_1<t_2+2\tau_0\\ (t_1-\tau_0,s_0] & \text{if } t_2-2\tau_0<t_1<t_2\\ (t_1-\tau_0,t_1+\tau_0) & \text{if } |t_1-t_2|\geq 2\tau_0\end{cases} \tag{A.29}$$

If $t_1\in\psi(s_1)$ then $q(s_1)=V(t_1)-\lambda(s_1-t_1)$ and $q(s_1)\geq V(t_2)-\lambda(s_1-t_2)$, so $s_1\in S(t_1,t_2)$. Similarly, $s_2\in S(t_2,t_1)$. Now suppose $t_1<t_2$. If $t_2-t_1<2\tau_0$ then, from (A.29), $S(t_1,t_2)=(t_1-\tau_0,s_0]$ and $S(t_2,t_1)=[s_0,t_2+\tau_0)$, so that $s_1\leq s_2$. If $t_2-t_1\geq 2\tau_0$ then, from (A.29), $S(t_1,t_2)=(t_1-\tau_0,t_1+\tau_0)$ and $S(t_2,t_1)=(t_2-\tau_0,t_2+\tau_0)$, so that $s_1<s_2$. Thus $t_1<t_2$ implies $s_1\leq s_2$. □

*Lemma A.*6. Suppose conditions V.1–V.2 and L.2–L.4 hold, and define $R(J;v)$ by (4.6) with $v\geq 0$. Let $0\leq a<T$. Then $R((a,T];v)$ is an interval,

$$R((a,T];v)=\langle\eta(a,v),T+\tau_0) \tag{A.30}$$

with $-\tau_0<\eta(a,v)\leq T+\tau_0$.

The angle bracket notation for intervals is defined in section 4. The interval may be null, in which case we set $\eta(a,v)=T+\tau_0$.

Proof. First we prove the lemma for the case $v=0$.

If $R((a,T])=\varnothing$ then we set $\eta(a,0)=T+\tau_0$, and if $R^*((a,T])=\varnothing$ then $\eta(a,0)=-\tau_0$. Having disposed of these cases, let $t_1\in R((a,T])$ and $t_2\in R^*((a,T])$. Note that $J^c=[0,a]$.

Suppose $s_1\in\psi(t_1)$. If $s_1\in[0,a]$ then we would have $q(t_1)=q(t_1;\{s_1\})\leq q(t_1;[0,a])$, whereas $t_1\in R((a,T])$ implies $q(t_1)=q(t_1;(a,T])>q(t_1;[0,a])$, and therefore s_1 must be in $(a,T]$. Suppose $s_2\in\psi(t_2)$. According to (A.27), $q(t_2)=q(t_2;[0,a])$, so by lemma A.4 we can choose s_2 to be in $[0,a]$. We now have $s_1>s_2$, and so $t_1\geq t_2$ by lemma A.5. Since $R((a,T])$ and $R^*((a,T])$ are disjoint, this means that $R((a,T])$ lies to the right of $R^*((a,T])$. By construction,

$$R((a,T])\cup R^*((a,T])=\{t\,|\,q(t)>-\infty\}=(-\tau_0,T+\tau_0),$$

so $R((a,T])$ and $R^*((a,T])$ must be contiguous intervals. Therefore, $R((a,T])$ must be of the form $(\eta(a,0),T+\tau_0)$ or $[\eta(a,0),T+\tau_0)$. This proves the lemma for $v=0$.

Now suppose $v>0$. The continuity properties V.1–V.2 are not affected if the function V is replaced by $V+v\cdot 1\{[0,a]\}$, so the result holds in this case also. □

Remarks. (a) An obvious symmetry argument extends lemma A.6 to $R([0,a);v)$, with $0 < a \le T$ and $v \ge 0$, giving

$$R([0,a);v) = (-\tau_0, \zeta(a,v)\rangle \tag{A.31}$$

where $-\tau_0 \le \zeta(a,v) < T + \tau_0$.

(b) Under the conditions of lemma A.6, it can easily be shown that $q(t;J)$ is continuous in *t*, provided that $q(t;J) > -\infty$. It then follows that $R((a,T];v)$ and $R([0,a);v)$ are, in fact, open intervals. Since that result is not needed here, we omit the proof that $q(t;J)$ is continuous.

Lemma A.7. Suppose $J_n \downarrow J$, where $J_n \subseteq [0,T]$, $n = 1,2,...$ Assume conditions V.1–V.2 and L.2. Then $R(J_n) \downarrow R(J)$, where $R(\cdot) = R(\cdot;0)$ is defined by (4.6).

Proof. Suppose first that $t \in \bigcap_n R(J_n)$. Then $q(t;J_n) > q(t;J_n^c)$ for each *n*, which means that we can find $s_n \in J_n$ such that

$$V(s_n) - \lambda(t - s_n) > V(s') - \lambda(t - s') \text{ for all } s' \in J_n^c. \tag{A.32}$$

Without loss we can choose the sequence $s_1, s_2,...$ such that

$$V(s_n) - \lambda(t - s_n) \ge V(s_{n-1}) - \lambda(t - s_{n-1}), \tag{A.33}$$

because if (A.33) could fail only if $s_{n-1} \in J_n$, and in that case we can choose s_n to be equal to s_{n-1}. The sequence is bounded, so a subsequence converges to $s \in J$. Then $V(s_n) \to V(s\pm) \le V(s)$, by conditions V.1–V.2. Because V is bounded, $\lambda(t - s_n)$ is bounded, and therefore $\lambda(t - s_n) \to \lambda(t - s)$ by condition L.2. It follows from (A.32)–(A.33) that

$$V(s) - \lambda(t - s) > V(s') - \lambda(t - s') \text{ for all } s' \in \bigcup_n J_n^c = J^c$$

which shows that $q(t;J) > q(t;J^c)$, and therefore $t \in \bigcap_n R(J_n)$ implies $t \in R(J)$.

Conversely, suppose $t \in R(J)$. From $J_n \supseteq J$ it follows that $q(t;J_n) \ge q(t;J)$ and $q(t;J_n^c) \le q(t;J^c)$, so that $t \in R(J)$ implies $t \in R(J_n)$.

Having shown inclusion both ways, we have the result $R(J) = \bigcap_n R(J_n)$. □

Proof of Lemma 4.1. First, note that $R(J;v) \subseteq R((a,T];v) \cap R([0,b);v)$. Suppose *t* belongs to $R((a,T];v)$ and $R([0,b);v)$. Then, from (4.6),

$$\max\{q(t;(a,b)), q(t;[b,T])\} > q(t;[0,a]) + v$$

and

$$\max\{q(t;[0,a]), q(t;(a,b))\} > q(t;[b,T]) + v.$$

If we suppose $q(t;(a,b)) \le q(t;[b,T])$ then these inequalities lead to a contradiction, but if $q(t;(a,b)) > q(t;[b,T])$ then we find $t \in R(J;v)$. Therefore

$$\begin{aligned} R(J;v) &= R((a,T];v) \cap R([0,b);v) \\ &= \langle \eta(a,v), T+\tau_0) \cap (-\tau_0, \zeta(b,v)\rangle, \end{aligned} \tag{A.34}$$

where the second line comes from (A.30) and (A.31). By construction, $\eta(a,v)$ and $\zeta(b,v)$ are in the interval $[-\tau_0, T+\tau_0]$, so (A.34) implies the result (4.7), with the understanding that $R(J;v)$ is empty if $\eta(a,v) > \zeta(b,v)$, and may be empty if $\eta(a,v) = \zeta(b,v)$.

Now for continuity. Choose $\varepsilon > 0$ small enough that $\eta(a,0)+\varepsilon \in R((a,T];0)$. Then

$$q(\eta(a,0)+\varepsilon; (a,T]) > q(\eta(a,0)+\varepsilon; [0,a]) + \delta$$

for all sufficiently small $\delta > 0$. This implies $\eta(a,0)+\varepsilon \in R((a,T];\delta)$, so

$$\eta(a,0)+\varepsilon > \eta(a,\delta).$$

But when $\delta > 0$ we have $R((a,T];\delta) \subseteq R((a,T];0)$, which implies $\eta(a,0) \le \eta(a,\delta)$. It follows that $|\eta(a,\delta) - \eta(a,0)| < \varepsilon$ for all sufficiently small $\delta > 0$, i.e., $\eta(a,v)$ is right continuous at $v = 0$.

This argument does not work if $\eta(a,0) = T+\tau_0$. But if $R((a,T];0) = \varnothing$ then clearly $R((a,T];\delta) = \varnothing$ for any $\delta > 0$, so that $\eta(a,\delta) = T+\tau_0$ also, and right continuity of $\eta(a,v)$ at $v = 0$ follows trivially.

Similarly, $\zeta(a,v)$ is right continuous at $v = 0$. □

Proof of Lemma 4.2. First we show that $\eta(a) \le \zeta(b)$. If $\eta(a) = -\tau_0$ or if $\zeta(b) = T+\tau_0$ this is trivial, so suppose that $t_1 \in (-\tau_0, \eta(a))$ and $t_2 \in (\zeta(b), T+\tau_0)$. As in the proof of lemma A.6, we can choose $s_1 \in \psi(t_1) \cap [0,a]$ and $s_2 \in \psi(t_2) \cap [b,T]$. Then $s_1 < s_2$, so lemma A.5 says $t_1 \le t_2$, which implies $\eta(a) \le \zeta(b)$.

Suppose that $\eta(a) < \zeta(b)$. If $\nu > 0$, then $R((a,T];v) \subseteq R((a,T];0)$, which implies $\eta(a,v) \ge \eta(a,0)$. Similarly, $\zeta(b,v) \le \zeta(b,0)$. But $\eta(a,v)$ and $\zeta(b,v)$ are right-continuous in v at $v = 0$, according to lemma 4.1, so for all sufficiently small v

$$\eta(a) \le \eta(a,v) < \zeta(b,v) \le \zeta(b).$$

Therefore, using the representation (4.9),

$$
\begin{aligned}
K(J;v)-K(J;0) &= \int_{\eta(a)}^{\eta(a,v)} dt\Big(\exp[q(t;J^c)]-\exp[q(t;J)-v]\Big) \\
&\quad + \int_{\zeta(b,v)}^{\zeta(b)} dt\Big(\exp[q(t;J^c)]-\exp[q(t;J)-v]\Big) \\
&\quad + \int_{\eta(a)}^{\zeta(b)} dt\, e^{q(t)}\,(e^{-v}-1)\,,
\end{aligned} \tag{A.35}
$$

where we have used the fact that $q(t;J)=q(t)$ when $t\in R(J;0)$. Call the first integral $I_1(v)$. The interval $(\eta(a,0),\eta(a,v))$ is in $R((a,T];0)$ but not in $R((a,T];v)$, so

$$q(t;J)-v\le q(t;J^c)\le q(t;J)$$

and therefore

$$
\begin{aligned}
0\le I_1(v) &\le \int_{\eta(a)}^{\eta(a,v)} dt\exp[q(t;J)]\,(1-e^{-v}) \\
&\le [\eta(a,v)-\eta(a)]\,e^c(1-e^{-v})
\end{aligned} \tag{A.36}
$$

where c is an upper bound on V. Continuity of $\eta(a,v)$ then implies $I_1(v)=o(v)$ as $v\to 0+$. The same applies to the second integral. The last term is obviously differentiable, so we get

$$\lim_{v\to 0+}\frac{1}{v}[K(J;v)-K(J;0)]=-\int_{\eta(a)}^{\zeta(b)} dt\, e^{q(t)}$$

which is the required result.

We still have the possibility $\eta(a)=\zeta(b)$. In this case $\eta(a,v)\ge\zeta(b,v)$ for all $v\ge 0$ and therefore, from (4.10), $K(J;v)=K(J;0)$. Then the right derivative at $v=0$ is zero, which is the required result because the integral $K_1(J)$ is also zero when $\eta(a)=\zeta(b)$. □

Proof of Lemma 4.3. Using the same argument as in the proof of lemma A.6, we find that $R([a,T])$ lies to the right of $R^*([a,T])$, and that $R^*([0,a))$ also lies to the right of $R^*([a,T])$ except possibly for one point in common. But

$$R([a,T])\cup R^*([a,T])=R^*([0,a))\cup R^*([a,T])=(-\tau_0,T+\tau_0)\,,$$

so $R([a,T])$ must be the same interval as $R^*([0,a))$, except possibly for the left end-point. We therefore have $R([a,T])=\langle\,\zeta(a),T+\tau_0)$. Similarly, $R([0,b])=(-\tau_0,\eta(b)\rangle$. Then, as in the proof of lemma 4.1,

$$R([a,b])=R([a,T])\cap R([0,b])=\langle\,\zeta(a),\eta(b)\rangle\,. \tag{A.37}$$

Let $J_n = (a-1/n, b+1/n)$, so that $J_n \downarrow [a,b]$. From lemma A.7, $R(J_n) \downarrow R([a,b])$, so that $\eta(a-1/n) \to \zeta(a)$ and $\zeta(b+1/n) \to \eta(b)$, and therefore

$$P_0(J_n) \to \int_{\zeta(a)}^{\eta(b)} dt\, e^{q(t)} \Big/ \int dt\, e^{q(t)} .$$

But $J_n \downarrow [a,b]$ implies $P_0(J_n) \downarrow P_0([a,b])$, and the result (4.14) follows. □

Proof of Lemma 4.4. From the proofs of lemmas 4.2 and 4.3,

$$a < b \Rightarrow \zeta(a) \le \eta(a) \le \zeta(b) \le \eta(b), \tag{A.38}$$

and from lemma A.7,

$$\begin{aligned} &a_n \uparrow a \Rightarrow \eta(a_n) \uparrow \zeta(a) \text{ and } \zeta(a_n) \uparrow \zeta(a) \\ &b_n \downarrow b \Rightarrow \zeta(b_n) \downarrow \eta(b) \text{ and } \eta(b_n) \downarrow \eta(b). \end{aligned} \tag{A.39}$$

Let Λ be a mapping from intervals of $[0,T]$ to intervals of $[-\tau_0, T+\tau_0]$ defined by

$$\begin{aligned} &\Lambda((a,b)) = (\eta(a), \zeta(b)), \qquad \Lambda((a,b]) = (\eta(a), \eta(b)], \\ &\Lambda([a,b)) = [\zeta(a), \zeta(b)), \qquad \Lambda([a,b]) = [\zeta(a), \eta(b)], \end{aligned} \tag{A.40}$$

where $a \le b$ for the closed interval and $a < b$ in the other cases. $\Lambda(J)$ differs from $R(J)$ only in the treatment of the end-points. According to (A.38) the end-points of non-empty intervals $\Lambda(J_n)$ have the same ordering as the end-points of the corresponding intervals J_n. It follows that

$$\begin{aligned} &J' \cap J = \varnothing \Rightarrow \Lambda(J') \cap \Lambda(J) = \varnothing, \\ &\quad J' \subset J \Rightarrow \Lambda(J') \subseteq \Lambda(J), \end{aligned} \tag{A.41}$$

and in particular

$$t \in J \Rightarrow \zeta(t) \in \Lambda(J) \text{ and } \eta(t) \in \Lambda(J). \tag{A.42}$$

Let $J = \langle a,b \rangle$, $J_k = \langle a_k, b_k \rangle$, $\Lambda(J) = \langle c,d \rangle$, and $\Lambda(J_k) = \langle c_k, d_k \rangle$. Under the conditions of the lemma, $\Lambda(J_1), \Lambda(J_2), \ldots$ is a sequence of disjoint intervals, and we want to show that $\bigcup_{n=1}^{\infty} \Lambda(J_n) = \Lambda(J)$. It follows at once from (A.41) that $\bigcup_{n=1}^{\infty} \Lambda(J_n) \subseteq \Lambda(J)$, so the problem is to show that $\Lambda(J) \subseteq \bigcup_{n=1}^{\infty} \Lambda(J_n)$.

Suppose $t \in \Lambda(J)$. If *t* is an end-point of $\Lambda(J)$ then, from (A.42), $t \in \Lambda(J_n)$ where J_n contains the corresponding end-point of *J*. This means we need only consider $t \in (c,d)$. Let

$$b^* = \sup_n \{ b_n \mid t' < t \ \text{ for all } \ t' \in \Lambda(J_n) \},$$
$$a^* = \inf_n \{ a_n \mid t' > t \ \text{ for all } \ t' \in \Lambda(J_n) \}. \tag{A.43}$$

From (A.39), $\zeta(b^*) \le t \le \eta(a^*)$, and (A.38) then implies $b^* \le a^*$. In the special case $b^* = -\infty$ we set $\zeta(b^*) = -\infty$, and if $a^* = \infty$ we set $\eta(a^*) = \infty$. By construction $\eta(a^*) > c$, and therefore $a^* \ge a$, with strict inequality except when $a \in J$. Similarly $b^* \le b$, with strict inequality except when $b \in J$, and therefore $J \cap [b^*, a^*]$ is not empty.

First, suppose $s \in J$ such that $b^* < s < a^*$. Let $s \in J_n$. Then $b^* < b_n$, and it follows from the definition of b^* that $t \le d_n$, with strict inequality except when $d_n \in \Lambda(J_n)$. Similarly, $a^* > a$ implies $t \ge c_n$, with strict inequality except when $c_n \in \Lambda(J_n)$. Therefore $t \in \Lambda(J_n)$ as required.

That argument does not work if $J \cap (b^*, a^*) = \varnothing$. Since $J \cap [b^*, a^*] \ne \varnothing$, three exceptional cases are possible: $a^* = b^* \in J$, $a^* = a \in J$, or $b^* = b \in J$. Suppose $a^* = b^* \in J$, and let $a^* \in J_n$. Then $\eta(a^*) \in \Lambda(J_n)$ and $\zeta(b^*) \in \Lambda(J_n)$ by (A.42). But $\zeta(b^*) \le t \le \eta(a^*)$, and therefore $t \in \Lambda(J_n)$. Next suppose $a^* = a \in J$, and let $a^* \in J_n$. In this case $c = \zeta(a)$, so by (A.42) we have $c \in \Lambda(J_n)$ and $\eta(a^*) \in \Lambda(J_n)$. But $c < t \le \eta(a^*)$, so $t \in \Lambda(J_n)$ in this case also. A similar argument applies to the case $b^* = b$.

This shows that $t \in \Lambda(J)$ implies $t \in \Lambda(J_n)$ for some *n*, which we needed to prove that $\Lambda(J) = \bigcup_n \Lambda(J_n)$. Because the intervals $\Lambda(J_n)$ are disjoint, additivity of Lebesgue measure then gives

$$\int dt\, 1\{t \in \Lambda(J)\} e^{q(t)} = \sum_n \int dt\, 1\{t \in \Lambda(J_n)\} e^{q(t)}, \tag{A.44}$$

and the result (4.18) follows at once. □

Proof of Lemma 5.1. Suppose $\eta(t) > (t)$, and let $s \in [\zeta(t), \eta(t)]$ with $s \ne t$. Then $q(s; \{t\}) > q(s; \{t\}^c)$, and therefore $V(t') - \lambda(s - t')$ is maximized at $t' = t$. Suppose first that *V* is differentiable at *t*. Then a necessary condition for a maximum is

$$V'(t) + \lambda'(s - t) = 0 \quad (s \ne t). \tag{A.45}$$

Since $\lambda(\tau)$ is strictly increasing, there is at most one *s* satisfying (A.45). This contradicts the assumption that $[\zeta(t), \eta(t)]$ is a non-degenerate interval, so *V* cannot be differentiable at *t*.

Next, suppose that V is not differentiable at t, but has unequal left and right derivatives $V'(t-)$ and $V'(t+)$. The necessary conditions for there to be a local maximum of $V(t') - \lambda(s - t')$ at $t' = t$ are now

$$V'(t+) \le -\lambda'(s-t) \quad \text{and} \quad V'(t-) \ge -\lambda'(s-t) \quad (s \ne t)$$

This is obviously not possible unless $V'(t+) < V'(t-)$ and either $V'(t+) \le -\lambda'(0+)$ or $V'(t-) \ge \lambda'(0+)$. □

Proof of Lemma 5.2. From condition D.1 and lemma 5.1 we have $\zeta(t) = \eta(t) \equiv \eta$. From condition D.4 we have $\zeta(t - dt) < \eta < \eta(t + dt)$ for all sufficiently small $dt > 0$, so that $\eta \in R([t - dt, t + dt])$ and therefore $q(\eta) = q(\eta; [t - dt, t + dt])$. Then, from lemma A.4, $q(\eta) = V(s) - \lambda(\eta - s)$ for some point s such that $|s - t| \le dt$. Letting $dt \to 0$ and using the continuity properties of V and λ, we get $q(\eta) = V(t) - \lambda(\eta - t)$. According to the definition of q, this means that $V(s) - \lambda(\eta - s)$ is maximized at $s = t$.

The necessary condition for a local maximum depends on whether or not $\eta = t$, because $\lambda(\tau)$ need not be differentiable at $\tau = 0$. If $\eta \ne t$ then the first-order condition is

$$V'(t) + \lambda'(\eta - t) = 0, \tag{A.46}$$

which is possible only if $|V'(t)| \ge \lambda'(0+)$. If $|V'(t)| \le \lambda'(0+)$, then the condition for a local maximum is $\eta = t$.

If $|V'(t)| < \lambda'(0+)$, then by condition V.3 this inequality holds in a neighborhood of t. We can therefore differentiate (A.46) to get

$$V''(t) + \lambda''(\eta - t)\,(d\eta/dt - 1) = 0,$$

which is the same as (5.1). Similarly, if $|V'(t)| > \lambda'(0+)$ then $\eta(t) = t$ in a neighborhood of t, so that $d\eta/dt = 1$. □

Proof of Lemma 7.1. Suppose $s_1 < s_2$, and let $r = r(s_1, s_2, \theta)$ be the right-hand boundary of the interval $S(s_1, s_2)$ defined by (A.28). From the continuity properties of V and λ,

$$V(s_1, \theta) - \lambda(r - s_1) = V(s_2, \theta) - \lambda(r - s_2). \tag{A.47}$$

From (A.27) and (3.10)–(3.11),

$$\begin{aligned} R^*((t_0, T]) &= \{t \mid t \in S(s_1, s_2) \text{ for some } s_1 \in [0, t_0] \text{ and all } s_2 \in (t_0, T]\} \\ &= \{t \mid t \le \inf_{s_2 \in (t_0, T]} r(s_1, s_2, \theta) \text{ and } t > s_1 - \tau_0 \text{ for some } s_1 \in [0, t_0]\} \end{aligned} \tag{A.48}$$

and therefore

$$\eta(t_0) = \sup_{s_1 \in [0, t_0]} \Big(\inf_{s_2 \in (t_0, T]} r(s_1, s_2, \theta) \Big) . \tag{A.49}$$

Put $\delta r = r(s_1, s_2, \theta + \delta\theta) - r(s_1, s_2, \theta)$. If $\delta r \leq (s_2 - s_1)$ then, because λ is strictly convex,

$$\lambda(r + \delta r - s_1) - \lambda(r - s_1) > \delta r \, \lambda'(r - s_1)$$

and

$$\lambda(r + \delta r - s_2) - \lambda(r - s_2) \leq \frac{\delta r}{s_2 - s_1} \big(\lambda(r - s_1) - \lambda(r - s_2) \big).$$

If $\delta r \geq (s_2 - s_1)$ then instead

$$\lambda(r + \delta r - s_1) - \lambda(r + \delta r - s_2) > (s_2 - s_1) \, \lambda'(r + \delta r - s_2)$$

and

$$\lambda(r - s_1) - \lambda(r - s_2) \leq \frac{s_2 - s_1}{\delta r} \big(\lambda(r + \delta r - s_2) - \lambda(r - s_2) \big).$$

In either case, therefore,

$$[\lambda(r + \delta r - s_1) - \lambda(r + \delta r - s_2)] - [\lambda(r - s_1) - \lambda(r - s_2)] > \frac{c}{2} \delta r \, (s_2 - s_1)$$

where $c > 0$ is a lower bound on $\lambda''(\tau)$. Substituting from (A.47) and (7.2), and using the corresponding result for $\delta r < 0$, we get

$$\big| r(s_1, s_2, \theta + \delta\theta) - r(s_1, s_2, \theta) \big| < \frac{2}{c} \big| \Delta(s_2, s_1, \theta + \delta\theta) - \Delta(s_2, s_1, \theta) \big| .$$

By assumption Δ is continuous in θ uniformly in s_1 and s_2 for $0 \leq s_1 \leq t_0 \leq s_2 \leq T$, and therefore so is *r*.

Applying this result in (A.49), it follows that $\eta(t_0)$ is continuous in θ. A similar argument shows that $\zeta(t_0)$ is continuous. □

References

Ben-Akiva, M. and T. Watanatada (1981). "A continuous spatial choice logit model," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. C. F. Manski and D. McFadden, MIT Press.

Billingsley, P. (1986). *Probability and Measure*, 2nd edition, Wiley.

Dagsvik, J. K. (1988). "The continuous generalized extreme value model with special reference to static models of labor supply," Discussion Paper No. 31, Central Bureau of Statistics, Oslo.

de Haan, L. (1984). "A spectral representation for max-stable processes," *Annals of Probability*, 12, pp. 1194–1204.

Huber, P. J. (1965). "The behavior of maximum likelihood estimates under non-standard conditions," in *Proceedings of the Fifth Berkeley Symposium on Statistics and Probability*, Vol. 1, ed. J. Neyman, University of California Press.

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

McFadden, D. (1981). "Econometric models of probabilistic choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. C. F. Manski and D. McFadden, MIT Press.

McFadden, D. (1984). "Qualitative response models," in *Handbook of Econometrics*, Vol. 2, ed. Z. Griliches and M. D. Intriligator, North-Holland.

Small, K. (1987). "A discrete choice model for ordered alternatives," *Econometrica*, 55, pp. 409–424.