COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 830

SEMIPARAMETRIC ESTIMATION OF MONOTONIC AND CONCAVE UTILITY FUNCTIONS: THE DISCRETE CHOICE CASE

Rosa L. Matzkin

April 1987

SEMIPARAMETRIC ESTIMATION

OF MONOTONIC AND CONCAVE UTILITY FUNCTIONS:

THE DISCRETE CHOICE CASE

By Rosa L. Matzkin[1]

Cowles Foundation for Research in Economics
Yale University
New Haven, CT 06520

First version: June 1986
Latest version: March 1987

1 An earlier version of this paper was part of my Ph.D Thesis, and was presented at the 1986 summer meetings of the Econometric Society under the title "Nonparametric Estimation of Utility Functions for Discrete Choice Models." I am indebted to Professors Lung-Fei Lee, Marcel K. Richter and Christopher Sims for their valuable advice. I wish to thank Philip Dybvig, Vassilis Hajivassiliou, and Alvin Klevorick for their comments.

ABSTRACT

This paper develops a semiparametric method for estimating the nonrandom part V( ) of a random utility function $U(v,\omega) = V(v) + e(\omega)$ from data on discrete choice behavior. Here v and ω are, respectively, vectors of observable and unobservable attributes of an alternative, and $e(\omega)$ is the random part of the utility for that alternative. The method is semiparametric because it assumes that the distribution of the random parts is known up to a finite-dimensional parameter θ, while not requiring specification of a parametric form for V( ).

The nonstochastic part V( ) of the utility function U( ) is assumed to be Lipschitzian and to possess a set of properties, typically assumed for utility functions. The estimator of the pair (V,θ) is shown to be strongly consistent.

1. INTRODUCTION

In theoretical microeconomics, we represent preferences using preference orderings that, if they have a utility representation at all, satisfy only general properties such as monotonicity, concavity and continuity. By contrast, empirical estimation of preferences normally proceeds by specifying a parametric functional form with finitely many parameters. In this paper, we propose a maximum likelihood technique for the semiparametric estimation of utility functions that uses only the assumptions derived from the microeconomic theory of consumer behavior, thereby bridging the gap between the theoretical and empirical literatures.

The resulting estimator is strongly consistent, as I demonstrate by adapting Wald's proof of the consistency of the maximum likelihood estimator in abstract spaces (Wald (1949)). The adaptation substitutes the assumption that the space of utility functions be compact in the supremum norm for one of Wald's assumptions and specializes the result to our context. The required compactness is shown to be an implication of the theoretical assumptions and an additional Lipschitz condition.

Our analysis focuses on <u>probabilistic choice models</u>, of which the logit model is the most familiar example [McFadden (1974, 1976, 1981, 1984)]. In these models, the consumer faces a finite number of alternatives from which to choose. The utility of each alternative is given by an underlying nonstochastic <u>subutility function</u> plus unobservable randomness. The subutility function depends on a vector of observable attributes of the alternative. We are assumed to know the distribution of the vector of unobservable noise, up to a finite-dimensional parameter vector. Given sample data on attributes of alternatives and actual choices of individuals, we want to estimate both the subutility function and the parameter of the distribution of the random vector.

Instead of assuming that the subutility function belongs to a particular parametric class, we assume that it lies in the class of functions satisfying one of several sets of restrictions. The basic restrictions are that the function is monotonic, concave, and satisfies a Lipschitz condition. I also consider the classes of functions having these three properties and satisfying one of the following conditions: (a) homogeneity of degree one, (b) weak separability, or (c) additive separability. The class of functions satisfying each restriction is compact under the supremum norm, which is what our proof of consistency requires. Our estimate of the subutility function converges almost surely uniformly to the true function.

Although it may seem difficult to maximize the likelihood function over a general function space, we use methods in the spirit of Afriat (1967a, 1967b, 1972, 1973, 1981), Diewert (1973) and Varian (1982,1983) to reduce the maximization to a finite dimensional problem for each number of observations. (Of course, the dimensionality of the problem increases with the number of observations, or else we would not have convergence for all functions in an infinite dimensional space.) For any given value for the parameter of the distribution of the random terms, the likelihood function depends at most on the value and subgradient of the subutility function at finitely many points. Following Afriat, the various restrictions on the subutility function at these points can be represented using simple inequality constraints. Therefore, the maximization over a function space can be reduced to a finite dimensional maximization with linear inequality constraints.

The estimation methods proposed in Gallant (1981,1982) and Gallant and Golub (1984) could also be employed to estimate utility functions without assuming a particular parametric structure for them. These methods consist of employing Fourier flexible forms to obtain a characterization of the function in terms of an infinite vector of parameters [Gallant (1981)]. Homogeneity and concavity conditions are then imposed on the function to be estimated by restricting the values of the parameters of the Fourier approximations [Gallant (1982), Gallant and Golub (1984)]. These methods apply only to continuously differentiable functions, and have not yet been developed to impose monotonicity, not-linearly homogeneous separable functions, or other restrictions our method can encompass.

In probabilistic choice models, estimation of the utility function proceeds on the assumption that the nonrandom subutility function possesses a particular parametric structure. Manski (1975,1985), Han (1985a,1985b) and, in a maximum likelihood context, Cosslett (1983), have developed estimation methods that relax the distributional assumption on the random part of the utility function while still maintaining a parametric structure for the nonrandom part.

Manski (1975,1985) and Han (1985b) assume the nonrandom part of the utility function for each alternative to be linear in a finite number of parameters. The random part of the utility function is assumed to be distributed with an unknown distribution. Manski and Han develop strongly consistent estimators for the parameters of the nonrandom subutility. While Manski's method is applicable to polychotomous choices, Han's method applies only to dichotomous choices. On the other hand, and in contrast to Manski's estimator, Han's estimator has been shown to be asymptotically normally distributed.

Cosslett (1983) assumes the nonrandom part of the utility function to possess some known parametric functional form while the distribution of the random term of the utility is assumed to be unknown. Cosslett develops a distribution-free maximum likelihood method that provides a strongly consistent estimator of both the cumulative distribution function of the difference of the random parts of any observed pair of alternatives, and the parameter vector of the nonrandom subutility. Cosslett's method applies only to observations from dichotomous choices.

For classical consumer demand data (prices and quantities of commodities), Varian (1985) and Epstein and Yatchew (1985) have employed Afriat's inequalities to develop a statistical test for the maximization hypothesis or equivalently, for the existence of a utility function generating the demand observations. Their method, however, does not apply to random utility functions; in their models, random terms are implicitly assumed to be generated either by measurement or by optimization errors.

Section 2 contains three subsections. Subsection 2.1 specifies the model. Subsection 2.2 presents the random variables involved and their distributions, and subsection 2.3 presents the likelihood function and defines the maximum likelihood estimator. Section 3 presents the assumptions and the main result on the consistency of the estimator. Section 4 presents the method to compute the estimates. Finally, the Appendix contains Lemmas and proofs.

2. THE MODEL AND THE LIKELIHOOD FUNCTION

2.1. *The Model*

As usual in random utility models for discrete choices, assume that there exists a population M of consumers and a set $A = \{1, \ldots, J\}$ of alternatives. Each individual i in M must choose a single alternative from the alternatives set A. The alternatives may be simple, such as taking the car or the bus to work, or complex, such as a pair including both house location and a means of commuting. Consumer i's utility for alternative j in A is assumed to be additively separable into a function $V^0 : Z \to R$ of observable attributes in a subset Z of R^K_+, and a measurable function $e : \Omega \to R$ of unobservable attributes in Ω. Denote by z^i_j the observable attributes vector of alternative j for individual i and by e^i_j the value of e at the unobservable attributes vector ω^i_j of alternative j for individual i. For example, z^i_1 may be the time and cost incurred by individual i if he takes alternative $j = 1$ (car) to go to work, and ω^i_1 may be the comfort derived from alternative 1. The vector z^i_j may include socioeconomic characteristics of individual i, such as his income.

Each individual i in M is assumed to choose one alternative $j(i)$ in A by maximizing $V^0(\ ) + e(\ )$ over the alternatives set $\{1, \ldots, J\}$.[2] For each i, define $s^i = (s^i_1, \ldots, s^i_J)$ as follows: $s^i_k = 1$ if $k = j(i)$ and $s^i_k = 0$ otherwise $(k = 1, \ldots, J)$. Since $j(i)$ depends on the observable attributes $z^i_1, \ldots, z^i_J$ and on the unobservable attributes ω^i_1,

$\ldots, \omega_J^i$, we can write $(s_1^i, \ldots, s_J^i)$ as

$$(1) \qquad (s_1^i, \ldots, s_J^i) = \psi(z_1^i, \ldots, z_J^i, \omega_1^i, \ldots, \omega_J^i; V^0(\ ), e(\ ))$$

for some function ψ.

We want to estimate V^0, the common nonstochastic utility function, from observations of $s_1^i, \ldots, s_J^i$ and $z_1^i, \ldots, z_J^i$ for $i = 1, \ldots, n$.

2.2. *Random Variables and Probabilities*

The vector $z^i = (z_1^i, \ldots, z_J^i)$ will be assumed to be a random variable distributed across the population of consumers with some probability measure G determined by a density function $g : Z^J \to R$, where $Z \subset R_+^K$. For example, if z_1^i is time spent traveling by car, z_1^i will be partially determined by how close i lives to a highway. Similarly, the unobservable $\omega^i = (\omega_1^i, \ldots, \omega_J^i)$ will be assumed to be a random variable distributed with some probability measure Q characterized by a density function $q : \Omega^J \to R$, which may depend on some parameter $\theta^0 \in \Theta \subseteq R^L$.

Hence, by (1), $(s_1^i, \ldots, s_J^i)$ is a random variable whose probability measure depends on G and $Q(\theta^0)$. In particular, the conditional probability of s^i given $z_1^i, \ldots, z_J^i$ is given by

$$(2) \qquad P(s = s^i | z^i; V^0, \theta^0) = \Pi_{j=1}^J [P_j^i(V^0, \theta^0)]^{s_j^i}$$

where $P_j^i(V^0, \theta^0)$ denotes the probability (under $Q(\theta^0)$) that individual

i will choose alternative j when $z^i = (z_1^i, \ldots, z_J^i)$ is the vector of individual i's observable attributes for alternatives $1, \ldots, J$, and V^0 is the common non-random part of the utility function $U(\ )$ of each individual i.

By our maximization assumption, it follows that, if the distribution of $(e_1^i, \ldots, e_J^i)$ is continuous,

$$(3) \quad P_j^i(V^0, \theta^0) = \text{Prob}\{V^0(z_j^i) + e(\omega_j^i) > V^0(z_k^i) + e(\omega_k^i),\ k = 1, \ldots, J;\ k \neq j\}$$

$$= Q(\theta^0)(\{(\omega_1^i, \ldots, \omega_J^i) | V^0(z_j^i) + e(\omega_j^i) > V^0(z_k^i) + e(\omega_k^i),\ k=1, \ldots, J; k \neq j\}).$$ [3]

Since both s^i and z^i are random variables (s^i depends on G and $Q(\theta^0)$, and z^i depends on G), the vector $(s_1^i, \ldots, s_J^i, z_1^i, \ldots, z_J^i)$ is a random vector whose value depends on G and $Q(\theta^0)$. Then

$$(s_1^i, \ldots, s_J^i, z_1^i, \ldots, z_J^i) = \chi(z_1^i, \ldots, z_J^i, \omega_1^i, \ldots, \omega_J^i; V^0(\ ), e(\ ))$$

for some function χ. For each i, we will denote this vector by x^i. Let P_0 denote its probability measure. The joint probability-density $f(\ )$ of P_0 with which x^i is distributed can be easily obtained from $g(\ )$ and $P(\ )$ since

$$f(x^i; V^0, \theta^0) = f(s^i, z^i; V^0, \theta^0)$$

$$(4) \qquad = g(z^i) \ P(s = s^i \mid z^i; V^0, \theta^0)$$

$$= g(z^i) \ \Pi_{j=1}^J [P_j^i(V^0, \theta^0)]^{s_j^i}$$

2.3. *Likelihood Function*

For any n observations $x^1, \ldots, x^n$ on n individuals, we will denote by $x^{(n)}$ the ordered n-tuple $(x^1, \ldots, x^n)$. If we let V^0 and θ^0 vary across different functions V in a set W and parameter values θ in a set Θ, we will obtain different density functions $f(x;V,\theta)$. We will denote by $L(x^{(n)}, V, \theta)$ the joint probability-density of $(x^1, \ldots, x^n)$ under $f(x;V,\theta)$ so that $L(x^{(n)},V,\theta)$ is a "likelihood function."

We will assume that any n observations $(x^1,..,x^n)$ are independent.[4] Hence, for each V and θ

(5) $$L(x^{(n)}, V, \theta) = \Pi_{i=1}^{n} \; f(x^i; V, \theta)$$

In Section 4 we will show how to find a maximum likelihood estimate of the pair (V^0,θ^0), a function $V_n : Z \to R$ and a parameter value $\theta_n \in R^L$ that will maximize the likelihood function (5) over a predetermined set of functions W, and a predetermined parameter space Θ. In the next Section we show that, under quite general conditions, these estimators are strongly consistent: as n tends to infinity, V_n converges almost surely uniformly to V^0, and θ_n converges almost surely to θ^0.

3. CONSISTENCY

Denote the set $\{s=(s_1, \ldots, s_J) | s_j \in \{0,1\}, j = 1, \ldots, J, \Sigma_{j=1}^{J} s_j = 1\}$ by S and the set $S \times Z^J$ by X. Hence, $s \in S$ and $x = (s, z) \in X$. Z will be assumed to be a convex and compact subset of R_+^K and, for normalization purposes, it will be assumed that $0 \in Z$.

Measure the distance $d(V, V')$ between any two functions $V : Z \to R$ and $V' : Z \to R$ by $d(V, V') = \sup\{|V(z) - V'(z)| \mid z \in Z\}$.

Let W be a compact set (with respect to d) of real-valued, continuous functions on Z , which contains V^0 . And let Θ be a compact set (with respect to the Euclidean metric $\delta = \|\cdot\|$,) which contains θ^0. Define the metric $r : (W \times \Theta) \times (W \times \Theta) \to R_+$ by

$$r\,[\,(V,\theta)\,,\,(V',\theta')\,] = d(V,V') + \delta(\theta,\theta').$$

It is easy to verify that $[\,(W \times \Theta)\,,\, r\,]$ is a compact metric space.

For each $z = (z_1,..,z_J) \in Z^J$, we will denote $(V(z_1), \ldots, V(z_J))$ by $(V_1, \ldots, V_J)$; and for each $x = (s,z) \in S \times Z^J$, we will denote the probability of observing s when z is the vector of the attributes of the alternatives by $P(s,z,V,\theta)$. Hence,

$$\begin{aligned} P(x,V,\theta) &= P(s,\, z,\, V(z_1),\, \ldots,\, V(z_J),\, \theta) \\ &= P(s,\, z,\, V_1,\, \ldots,\, V_J,\theta) \\ &= \Pi_{j=1}^{J}\, [P(s_j = 1|z;\, V_1,\ldots,\, V_J,\, \theta]^{s_j} . \end{aligned}$$

We will show the consistency of the maximum likelihood estimator of (V^0,θ^0) under the above assumptions on Z, $(W \times \Theta)$, and the following conditions on $g(z)$ and $P(x,V,\theta)$:

ASSUMPTION G.1: *$g(\ )$ is measurable, and its support is Z.*

ASSUMPTION G.2: *$g(\ )$ is uniformly bounded.*

ASSUMPTION G.3: *For all z in Z,* $g(z) > 0$.

ASSUMPTION P.1: *For all $(s,z) \in X$, and all $(V,\theta) \in (W \times \Theta)$, $P(s,z,V,\theta) > 0$.*

ASSUMPTION P.2: *For all $(s,z) \in X$ $P(x,V,\theta)=P(s,z,V_1,..,V_J,\theta)$ is continuous at all $(V_1,..,V_J,\theta) \in (R_+^J \times \Theta)$.*

ASSUMPTION P.3: *For all $s \in S$ there exist $k,k' \in \{1,..,J\}$ such that for all $z \in Z^J$, $P(s,z,V,\theta^0) = P(s,z,V_1, \ldots,V_J,\theta^0)$ is strictly increasing in V_k and strictly decreasing in $V_{k'}$.*

ASSUMPTION P.4: *If $\theta \in \Theta$, $\theta \neq \theta^0$, and $V \in W$, either there exists $s' \in S$ such that for all $z \in Z^J$, $P(s',z,V,\theta) > P(s',z,V,\theta^0)$, or there exists $s^* \in S$ such that for all $z \in Z^J$, $P(s^*,z,V,\theta) < P(s^*,z,V,\theta^0)$.*

The following theorem states that, under the above assumptions, any sequence $\{(V_n,\theta_n)\}$ of maximum likelihood estimates of (V^0,θ^0), will almost surely converge to (V^0,θ^0) :

THEOREM 1: *Suppose that*

(1.i) $V^0 : Z \to R$ *belongs to a compact (with respect to the supremum metric d) set* W *of continuous functions* $V : Z \to R$ *, and* θ^0 *belongs to a compact (with respect to the Euclidean metric* δ*) set* $\Theta \subset R^L$,

(1.ii) Z *is a convex and compact subset of* R^K_+, $0 \epsilon Z$,

(1.iii) $g(\ )$ *satisfies G.1-G.3, and*

(1.iv) $P(\ ,\ ,\ )$ *satisfies P.1 to P.4.*

If for each n , $(V_n, \theta_n) \epsilon (W \times \Theta)$ *maximize the likelihood of* n *independent observations* $x^{(n)}$ *on the set* $(W \times \Theta)$ *for* $n = 1, 2, \ldots$, *then*

$$P_0 \ \{ \lim_{n\to\infty} r \ [\ (V_n, \theta_n\ ),\ (V^0, \theta^0)\ ] = 0\ \} = 1\ .$$

The proof of this theorem is based upon Wald's (1949) theorem on the strong consistency of maximum likelihood estimators on abstract spaces. We show in Appendix A that the assumptions (1.i)-(1.iv) imply that our model satisfies certain properties. These properties are all but one of Wald's assumptions. We then show in Appendix B that those properties and the compactness (with respect to r) of $(W \times \Theta)$ imply the conclusion of the theorem. (See Kiefer and Wolfowitz (1959) for a similar theorem.)

We now give two popular examples that satisfy our assumptions P.1 to P.4.

EXAMPLE 1: The set of alternatives A possesses three elements; and the random vector (e^i_1, e^i_2, e^i_3) is distributed with a Generalized Extreme Value distribution of the form

$$\Pr\left((e_1^i, e_2^i, e_3^i) \le (\eta_1, \eta_2, \eta_3)\right) = \exp\left[-\left[y_1 + \left(y_2^{1/(1-\theta)} + y_3^{1/(1-\theta)}\right)^{(1-\theta)}\right]\right]$$

where $y_1 = \exp(-\eta_1)$, $y_2 = \exp(-\eta_2)$, $y_3 = \exp(-\eta_3)$, and $\theta \in [0, \zeta]$ for some $\zeta > 0$.

McFadden (1978) shows that the choice probabilities $P(s_j = 1 | z; V_1, V_2, V_3, \theta)$ derived from this distribution are

$$P(s_1 = 1 | z;\ V,\ \theta) = \frac{\exp(V_1)}{[\exp(V_1) + [\exp(V_2)^{1/(1-\theta)} + \exp(V_3)^{1/(1-\theta)}]^{(1-\theta)}]}$$

$$P(s_2 = 1 | z;\ V,\ \theta) = \frac{\exp(V_2)^{1/(1-\theta)} \quad [(\exp(V_2)^{1/(1-\theta)} + \exp(V_3)^{1/(1-\theta)})^{-\theta}]}{[\exp(V_1) + [\exp(V_2)^{1/(1-\theta)} + \exp(V_3)^{1/(1-\theta)}]^{(1-\theta)}]}$$

$$P(s_3 = 1 | z;\ V,\ \theta) = \frac{\exp(V_3)^{1/(1-\theta)} \quad [(\exp(V_2)^{1/(1-\theta)} + \exp(V_3)^{1/(1-\theta)})^{-\theta}]}{[\exp(V_1) + [\exp(V_2)^{1/(1-\theta)} + \exp(V_3)^{1/(1-\theta)}]^{(1-\theta)}]}$$

Clearly, $P(s, z, V, \theta)$ satisfies Assumptions P.1 and P.2. Assumption P.3 is satisfied because for all $z \in Z^J$ and all $\theta \in [0, \zeta]$

$$\frac{\partial P(s_1 = 1 | z; V, \theta)}{\partial V_1} > 0, \qquad \frac{\partial P(s_1 = 1 | z; V, \theta)}{\partial V_2} < 0,$$

$$\frac{\partial P(s_2 = 1 | z; V, \theta)}{\partial V_1} < 0, \qquad \frac{\partial P(s_2 = 1 | z; V, \theta)}{\partial V_2} < 0,$$

$$\frac{\partial\ P(s_3=1|z;V,\theta)}{\partial\ V_1} < 0, \quad \text{and} \quad \frac{\partial\ P(s_3=1|z;V,\theta)}{\partial\ V_3} > 0\ .$$

And Assumption P.4 is satisfied because for all $z \in Z^J$ and $V \in W$

$$\frac{\partial\ P(s_1=1|z;V,\theta)}{\partial\ \theta} > 0.$$

EXAMPLE 2: The set of alternatives A possesses J elements; and each e^i_j ($i \in M$; $j=1,..,J$) is distributed, independently of e^i_k $(k\neq j)$, with the Gumbel distribution.[5]

McFadden (1974) shows that then the choice probabilities are

$$P(s_j=1|z;V,\theta) = \frac{\exp(V(z_j))}{\Sigma^J_{k=1} \exp(V(z_k))} \qquad (j=1,..,J)$$

In this example, P.1 to P.3 are satisfied for any θ. And, since only the function V^0 is estimated, P.4 is irrelevant.

4. ESTIMATION

In this section, we describe various compact sets, W, of continuous functions for which computation of the corresponding maximum likelihood estimates is feasible. The computation proceeds by solving a linear-inequality constrained maximization problem, and by interpolating between the obtained values.[6]

We present, in Lemma 1, the constraint set for the case in which W consists of all monotonic, concave functions $V : Z \to R$ that possess uniformly bounded subgradients and for which $V(0) = 0$. In Lemmas 7, 8 and 9, in Appendix C, we present the corresponding constraints for sets of functions having these properties and satisfying, in addition, one of the following conditions: (a) homogeneity of degree one, (b) weak separability, and (c) additive separability.

First note that given any set of observations $x^{(n)} = (x^1,..,x^n) = ((s^1,z^1),\ldots,(s^n,z^n))$ and any parameter θ, the value of $L(x^{(n)}, V, \theta)$ depends only on $V(z_1^1)$, ..., $V(z_J^1)$, ..., $V(z_1^n)$, ..., $V(z_J^n)$. If V and V' are two functions in W such that for all $i = 1, \ldots, n$ and all $j = 1,\ldots,J$, $V(z_j^i) = V'(z_j^i)$, then $L(x^{(n)}, V, \theta) = L(x^{(n)}, V', \theta)$. Hence, finding a function V_n in W and a parameter value θ_n in Θ, given $x^{(n)}$, is equivalent to the following procedure. First find an $n \times J$-dimensional vector $\underset{\sim}{V}^* = (V_1^{1*},\ldots,V_J^{1*},\ldots,V_1^{n*},\ldots,V_J^{n*})$ and a parameter value θ^* in Θ that maximize

$$(6) \qquad L(x^{(n)}, \underset{\sim}{V}, \theta) = \Pi_{i=1}^n g(z^i) \;\; \Pi_{j=1}^J [P(s_j^i = 1|z^i;\, V_1^i, \ldots, V_J^i,\theta)]^{s_j^i}$$

over the set of all vectors $\underset{\sim}{V} = (V_1^1, \ldots, V_J^n)$ for which there exists a function V in W such that $V_j^i = V(z_j^i)$ $(j = 1, \ldots, J;\ i = 1, \ldots, n)$, and the set of all parameter values θ in Θ . Second, interpolate between the values of $\underset{\sim}{V}^*$ to obtain a function V_n in W .

Since, as the next lemma shows, the set of all vectors $\underset{\sim}{V} = (V_1^1, \ldots, V_J^n)$, for which there exist a function V in W such that

$V_j^i = V(z_j^i)$ $(j = 1, \ldots, J;\ i = 1, \ldots, n)$, can be characterized by a finite set of linear inequalities, the search for $\underset{\sim}{V}^*$ is computationally feasible.

LEMMA 1: *Let* $B \in R_+^K$ *be some known vector. Let* W *be the set of all monotonic and concave*[7] *functions* $V : Z \to R$ *satisfying (i)* $V(0) = 0$, *and (ii) for each* $v \in Z$, *there exists a subgradient*[8] $T(v) \in R_+^K$ *such that* $T(v) \le B$.

Let $D = \{z_1, \ldots, z_M\}$ *be any given finite subset of the interior of Z. Then the set of all vectors* $\underset{\sim}{V} = (V_1, \ldots, V_M)$ *for which there exists a function* V *in* W *with* $V_i = V(z_i)$ $(i = 1, \ldots, M)$ *is the set of all vectors* $\underset{\sim}{V} = (V_1, \ldots, V_M)$ *for which there exists a vector* $T = (T^0, T^1, \ldots, T^M) \in R^{K\times(M+1)}$ $(T^m = (T_1^m, \ldots, T_K^m) \in R^K$, $m = 0, 1, \ldots, M)$ *such that*

$$(7) \qquad 0 \le T^j \le B ,$$

and

$$(8) \qquad V_i - V_j - T^j(z_i - z_j) \le 0 ,$$

for all $i, j = 0, 1, \ldots, M$, *and for* $z_0 = 0$ *and* $V_0 = 0$.

The proof of this Lemma is given in Appendix C. In the proof it is shown that the function $V : Z \to R$ defined by

$$(9) \qquad V(v) = \min\{V_j + T^j(v - z_j) \,|\, j = 0, 1, \ldots, M\},$$

for all $v \in Z$, where V_j and T^j , $j = 0, 1, \ldots, M$ satisfy (7) and (8), is a function that belongs to W . Moreover, for all z_j in D, $V(z_j) = V_j$ and T^j is in the subdifferential of V.[9]

Hence, given any set of observations $x^{(n)}$, we can obtain a maximum likelihood estimate (V_n, θ_n) of (V^0, θ^0) , by finding first vectors θ^* , $\underset{\sim}{V}^* = (V_1^{1*}, \ldots, V_J^{n*})$, and $\underset{\sim}{T}^* = (T^{0*}, T_1^{1*}, \ldots, T_J^{n*})$ that maximize

$$L(x^{(n)}, \underset{\sim}{V}, \theta) = \Pi_{i=1}^{n} g(z^i) \quad \Pi_{j=1}^{J} [P(s_j^i = 1 | z^i; V_1^i, \ldots, V_J^i, \theta)]^{s_j^i}$$

over Θ, and the set of all vectors $\underset{\sim}{V} = (V_1^1, \ldots, V_J^n) \in R^{nJ}$ and $\underset{\sim}{T} = (T^0, T^1, \ldots, T^{nJ}) \in R^{K(nJ+1)}$ satisfying (7) and (8); and then, through interpolation, obtaining a function V_n that belongs to W by defining for all $v \in Z$

$$V_n(v) = \min \{ T^{0*}v, \min \{V_j^{i*} + T_j^{i*} (v - z_j^i) \mid i=1,..,n; j=1,..,J \} \}$$

In fact, we can obtain a monotonic and concave function V_n *defined on all of* R_+^K, by defining for all $v \in R_+^K$

$$V_n(v) = \min \{ T^{0*}v, \min \{V_j^{i*} + T_j^{i*} (v - z_j^i) \mid i=1,..,n; j=1,..,J \} \}$$

As noted before, we can obtain maximum likelihood estimators for functions V^0 belonging to various compact sets, for which the constraint sets are presented in Lemmas 7-9, in Appendix C. The proofs of these Lemmas follow arguments that are very similar to those given in the proof of Lemma 1; therefore, their proof is omitted.

We finally note that, since g( ) does not depend on either V^0 or θ^0, the maximum likelihood estimate of (V^0, θ^0) for any given observations $x^1, .., x^n$ is independent of the specific functional form of g( ); hence, g( ) need not be known.

5. CONCLUSION

We have presented a maximum likelihood method of estimating the nonrandom part $V^0(\ )$ of the random utility function

$$U(v,\omega) = V^0(v) + e(\omega)$$

from observations on choices between a finite number J of alternatives. The method consists in first estimating the values of $V^0(\ )$ at the observed realizations of v, and second to interpolating between those values to obtain an estimate of $V^0(\ )$. The assumptions made on $V^0(\ )$ restrict the values that any estimator of $V^0(\ )$ may attain at each v. These restrictions are imposed in the estimation by requiring that the values and subgradient of the estimator satisfy a system of linear inequalities.

The method assumes that the distribution of the random vector $(e(\omega_1), .., e(\omega_J))$ is known up to a finite-dimensional parameter vector θ^0. The value of θ^0 is also estimated.

We have shown that the estimators of V^0 and θ^0 are strongly consistent: The estimator of V^0 converges almost surely uniformly to V^0 and the estimator of θ^0 converges almost surely to θ^0, as the number of observations tends to infinity.

APPENDIX A

In this Appendix, we show that the probabilistic choice model described in Sections 2 and 3 satisfies certain properties (Lemmas 2-6,) which will be used in Appendix B to prove Theorem 1. We remind the reader that $[\ (W \times \Theta),\ r\ ]$ is a compact metric space; W is a set of continuous functions on a convex and compact set $Z \subset R_+^K$, $0 \in Z$; Θ is a compact set in R^L; the density function $g(z)$ on attribute vectors satisfies Assumptions G.1-G.3, and the choice probability $P(x,V,\theta)$ satisfies Assumptions P.1-P.4 in Section 3.

LEMMA 2: *If* $\{(V_i,\theta_i)\}_{i=1}^{\infty}$ *is a sequence in* $(W \times \Theta)$ *and for some* $(V,\theta) \in (W \times \Theta)$, $\lim_{i\to\infty} r[(V_i,\theta_i),(V,\theta)] = 0$, *then, for all* x *in* $S \times Z$, $\lim_{i\to\infty} f(x, V_i, \theta_i) = f(x, V, \theta)$.

PROOF: Since convergence of the sequence $\{V_i\}$ to V with respect to the metric $d(\ )$ implies pointwise convergence of $\{V_i(z)\}$ to $V(z)$ for all $z \in Z$, and since by assumption P.1, $f(\ )$ is continuous on $(R_+^J \times \Theta)$ at each $(V_1, \ldots, V_J, \theta) = (V(z_1), \ldots, V(z_J), \theta) \in (R_+^L \times \Theta)$; it follows that for all x in $S \times Z$, $\{f(x, V_i, \theta_i)\}$ converges to $f(x,V,\theta)$.

Q.E.D.

LEMMA 3: *If* $(V,\theta) \in (W \times \Theta)$ *and* $(V,\theta) \neq (V^0,\theta^0)$ *then for some set* $A \subseteq X$ $P_0(A) > 0$, *and* $\int_A g(z)\, P(x,V,\theta)\, dx \neq \int_A g(z)\, P(x,V^0,\theta^0)\, dx$.

PROOF: We will first show that if $V \in W$ and $V \neq V^0$ then for any given $s \in S$,

(10) $\exists\ B_1 \subseteq Z^J$ such that for all $z \in B_1$ $P(s,z,V,\theta^0) > P(s,z,V^0,\theta^0)$; and

(11) $\exists\ B_2 \subseteq Z^J$ such that for all $z \in B_2$ $P(s,z,V,\theta^0) > P(s,z,V^0,\theta^0)$.

Since $V \neq V^0$ there exists $v \in Z$ for which $V(v) \neq V^0(v)$.

Suppose first that $V(v) > V^0(v)$. Since V and V^0 are continuous, there exists a neighborhood N in the interior of Z, containing v and such that for all w in N, $V(w) > V^0(w)$. Since $V(0)=V^0(0)=0$ and by assumption P.3 there exist $k,k' \in \{1,..,J\}$ such that, for all $z \in Z^J$, $P(s,z,V,\theta^0)$ is strictly increasing in V_k and strictly decreasing in $V_{k'}$, it follows that for all $z=(z_1,..,z_J)$ for which $z_k \in N$ and $z_j=0$ for $j=1,..,J$, $j\neq k$,

(12) $$P(x,V(z_1),V(z_2),...,V(z_J),\theta^0) > P(x,V^0(z_1),V^0(z_2),...,V^0(z_J),\theta^0),$$

and for all $z = (z_1,..,z_J)$ for which $z_{k'} \in N$ and $z_j=0$ for $j=1,..,J$, $j\neq k'$,

(13) $$P(x,V(z_1),V(z_2),...,V(z_J),\theta^0) < P(x,V^0(z_1),V^0(z_2),...,V^0(z_J),\theta^0) .$$

Since V and V^0 are continuous at $w=0$ and by Assumption P.2 $P(\ )$ is continuous at $(V_1,..,V_J) \in R^J_+$, it follows that there exists a neighborhood N^0 in the interior of Z such that for all $z=(z_1,..,z_J)$ for which $z_j \in N^0$ $(j=1,..,J;\ j\neq k)$ and $z_k \in N$, (12) holds; and, for all $z=(z_1,..,z_J)$ for which $z_j \in N^0$ $(j=1,..,J;\ j\neq k')$ and $z_{k'} \in N$, (13) holds.

Let $B_1 = \Pi^J_{j=1} N_j$, where $N_j = N^0$ $(j=1,..,J;j\neq k)$, and $N_k = N$; and let $B_2 = \Pi^J_{j=1} N_j$, where $N_j = N^0$ $(j=1,..,J;j\neq k')$, and $N_{k'} = N$. Then, (10) and (11) are satisfied.

Suppose now that $V(v) < V^0(v)$, then, by following the same arguments as in the previous case, it follows from Assumption P.3 that for all $z=(z_1,..,z_J)$ for which $z_{k'} \epsilon N$ and $z_j=0$ for $j=1,..,J$, $j \neq k'$, (12) holds; and, for all $z=(z_1,..,z_J)$ for which $z_k \epsilon N$ and $z_j=0$ for $j=1,..,J$, $j \neq k$, (13) holds. Hence, (10) holds for the set $B_1 = \Pi_{j=1}^{J} N_j$, where $N_j = N^0$ $(j=1,..,J;j \neq k')$, and $N_{k'} = N$; and (11) holds for the set $B_2 = \Pi_{j=1}^{J} N_j$, where $N_j = N^0$ $(j=1,..,J;j \neq k)$, and $N_k = N$.

We will now employ this result and Assumption P.4 to prove the Lemma.

Suppose that $(V,\theta) \epsilon (W \times \Theta)$, and $(V,\theta) \neq (V^0,\theta^0)$. By Assumption P.4 it follows that either

(14) $\exists\ s' \epsilon S$ such that for all $z \epsilon Z^J$ $P(s',z,V,\theta) \geq P(s',z,V,\theta^0)$, or

(15) $\exists\ s^* \epsilon S$ such that for all $z \epsilon Z^J$ $P(s^*,z,V,\theta) \geq P(s^*,z,V,\theta^0)$;

the corresponding inequality being strict whenever $\theta \neq \theta^0$.

If (14) is true it follows from (10) that for all $z \epsilon B_1$

(16) $P(s',z,V,\theta) \geq P(s',z,V,\theta^0) \geq P(s',z,V^0,\theta^0)$;

and if (15) is true it follows from (11) that for all $z \epsilon B_2$

(17) $P(s^*,z,V,\theta) \leq P(s^*,z,V,\theta^0) \leq P(s^*,z,V^0,\theta^0)$;

the second inequalities being strict whenever $V \neq V^0$.

Let $A = \{ (s,z) \epsilon X \mid s=s',\ z \epsilon B_1 \}$ if (14) is true; and let $A = \{ (s,z) \epsilon X \mid s=s^*,\ z \epsilon B_2 \}$ if (15) is true. Assumptions G.3 and P.1

imply that in either case $P_0(A) > 0$, and from (16), (17), and the fact that $(V, \theta) \neq (V^0, \theta^0)$ it follows that

$$\int_A g(z)\ P(x,V,\theta)\ dx \neq \int_A g(z)\ P(x,V^0,\theta^0)\ dx.$$

Q.E.D.

LEMMA 4: $\int_X |log\ f(x;\ V^0,\ \theta^0)|\ dP_0(x) < \infty$.

PROOF: By Assumption G.2 it follows that for some $K \in R$ and all $x \in X$, $|\ g(z)\ P(x,V^0,\theta^0)\ | < K$. Hence,

$$\int_X |\log\ f(x;\ V^0,\ \theta^0)|dP_0\ (x)$$

$$\leq \int_X |\log\ g(z)|g(z)P(x,V^0,\theta^0)\ dx + \int_X |\log\ P(x,V^0,\theta^0)|g(z)P(x,V,\theta^0)dx$$

$$\leq \int_X|\log\ g(z)|g(z)dx + K \int_X|\log\ P(x,V^0,\theta^0)|P(x,V^0,\theta^0)dx\ .$$

Since the ranges of g( ) and P( ) are included in an interval [0,K'], for some $K' > 0$, and the function $h(y) = y|\log y|$ has a bounded range on that interval, both integrals are bounded. Hence

$$\int_X|\log\ f(x;\ V^0,\theta^0)|\ dP_0(x) < \infty .$$

Q.E.D.

Following Wald (1949) we define the functions

$f' : (X \times W \times \Theta \times R_{++}) \to R$, and $f^{*} : (X \times W \times \Theta \times R_{++}) \to R$ by

$$(18) \quad f'(x,V,\theta,\varepsilon) = \sup_{(V',\theta')\epsilon(W\times\Theta)} \{f(x,V',\theta') | r[(V,\theta),(V',\theta')] < \varepsilon\},$$

and

$$(19) \quad f^*(x,V,\theta,\varepsilon) = \begin{cases} f'(x,V,\theta,\varepsilon) & \text{if } f'(x,V,\theta,\varepsilon) \geq 1 \\ 1 & \text{otherwise.} \end{cases}$$

LEMMA 5: *For any (V,θ) in $(W \times \Theta)$ and for all $\varepsilon > 0$, $f'(x,V,\theta,\varepsilon)$ is a measurable function of x .*

PROOF: Since $[(W \times \Theta), r]$ is a compact metric space, there exists a countable dense subset $C = \{(V_1,\theta_1), (V_2,\theta_2), \ldots\}$ of $(W \times \Theta)$. Hence, for any $x \epsilon X$ and $(V,\theta) \epsilon (W \times \Theta)$

$$(20) \quad \sup\{f(x, V', \theta') \mid r[(V',\theta'),(V,\theta)] < \varepsilon, (V',\theta') \epsilon (W \times \Theta) \}$$

$$= \sup\{f(x, V_i,\theta_i) \mid r[(V_i,\theta_i),(V,\theta)]) < \varepsilon, i = 1, 2, \ldots\} .$$

To see that (20) is true note that, since $C \subset (W \times \Theta)$,

$$\sup\{f(x, V_i,\theta_i) \mid r[(V_i,\theta_i),(V,\theta)]) < \varepsilon, i = 1, 2, \ldots\}$$

$$\leq \sup\{f(x, V', \theta) \mid r[(V',\theta'),(V,\theta)] < \varepsilon, (V',\theta') \epsilon (W \times \Theta) \}.$$

If the inequality were strict, that would imply that for some $(V',\theta') \epsilon (W \times \Theta)$

$$(21) \quad f(x,V',\theta') > \sup\{f(x,V_i,\theta_i)|r[(V_i,\theta_i),(V,\theta)]<\varepsilon, i=1,2,\ldots\} \geq f(x,V_i,\theta_i)$$

for i=1,2,... . Since $(V',\theta') \epsilon (W \times \Theta)$, and C is dense in $(W \times \Theta)$, there must exist a sequence $\{(V_k,\theta_k)\}_{k=1}^{\infty}$ in C such that $r[(V_k,\theta_k),(V',\theta')] \to 0$ as $k \to \infty$. But then, by assumption P.2, $f(x,V_k,\theta_k) \to f(x,V',\theta')$ contradicting (21). So, (20) holds.

From (20) it follows that for any $a \epsilon R$,

$$
\begin{aligned}
(22)\quad & \{x \epsilon X | \sup\{f(x,V',\theta') \mid r[(V',\theta'),(V,\theta)] < \varepsilon,\ (V',\theta') \epsilon (W \times \Theta)\ \} > a\} \\
&= \{x \epsilon X | \sup\{f(x,V_i,\theta_i) \mid r[(V_i,\theta_i),(V,\theta)] < \varepsilon;\ i=1,2,\ \ldots\} > a\}\ . \\
&= \cup_i \{x \epsilon X \mid f(x,\ V_i,\ \theta_i) > a\}\ ,
\end{aligned}
$$

where the union is over all i's for which $r[(V_i,\theta_i),(V,\theta)] < \varepsilon$ and $(V_i,\theta_i) \epsilon C$.

Since the V_i functions are measurable on $v \epsilon Z$, it follows from P.2 and the fact that S is a finite set that $\{x \epsilon X \mid f(x,\ V_i,\ \theta_i) > a\}$ is a measurable set, then $\cup_i \{x \epsilon X \mid f(x,\ V_i\ ,\ \theta_i\ ) > a\}$ is a measurable set. Then, by (22), $f(x,V,\theta,\varepsilon)$ is measurable on X .

Q.E.D.

LEMMA 6: *For any $(V,\theta) \epsilon (W \times \Theta)$ and for sufficiently small $\varepsilon > 0$ $\int_X \log f^*(x,V,\theta,e)\, dP_0(x)$ is finite.*

PROOF: By Assumption G.2 it follows that for some $K > 0$, all $x \epsilon X$, and all (V,θ) in $(W \times \Theta)$, $g(z)\ P(x,V,\theta) \le K$. Hence, $f'(x,V,\theta,\varepsilon) \le K$, and then $f^*(x,V,\varepsilon) = 1$ if $K < 1$, and $1 \le f^*(x,V,\varepsilon) \le K$ if $K > 1$.

The Lemma obviously holds if $K < 1$.

If $K > 1$ then, since by Lemma 5 $f'(x,V,\theta,\varepsilon)$ is measurable in x, the disjoint sets $C=\{x \in X \mid f'(x,V,\theta,\varepsilon) \leq 1\}$ and $D=\{x \in X \mid f'(x,V,\theta,\varepsilon) > 1\}$, are measurable. Hence

$$\int_X \log(f^*(x,V,\theta,\varepsilon))\, dP_0(x)$$

$$= \int_C \log(f^*(x,V,\theta,\varepsilon))\, dP_0\,(x) + \int_D \log(f^*(x,V,\theta,\varepsilon))\, dP_0(x)$$

$$= \int_D \log(f^*(x,V,\theta,\varepsilon))\, dP_0(x)$$

$$\leq \int_D \log(K)\, dP_0(x)$$

$$< \infty\ .$$

Q.E.D.

APPENDIX B

In this Appendix, we make use of Lemmas 2-6 in Appendix A and prove Theorem 1.

PROOF OF THEOREM 1: First we note that by Lemmas 6, 4, 3 and Lemma 1 in Wald (1949) it follows that for any (V,θ) in $(W \times \Theta)$ for which $(V,\theta) \neq (V^0,\theta^0)$,

$$\text{(23)} \qquad E \log f(X,V,\theta) < E \log f(X,V^0,\theta^0)\ .$$

And, by Lemmas 2, 6 and Lemma 2 in Wald (1949), for any (V,θ) in $(W \times \Theta)$,

$$\text{(24)} \qquad \lim_{\varepsilon\to 0}\ E \log f'(X,V,\theta,\varepsilon) = E \log f(X,V,\theta)\ ,$$

where the expectation is taken with respect to $P_0(\ )$.

From (23) and (24) it follows that for any (V,θ) in $(W \times \Theta)$ for which $(V,\theta) \neq (V^0,\theta^0)$, there exists $\varepsilon(V,\theta) > 0$ such that

$$(25) \qquad E \log f'(X, V, \theta, \varepsilon(V,\theta)) < E \log f(X, V^0, \theta^0) .$$

Let Y be any closed subset of $(W \times \Theta)$ which does not contain (V^0,θ^0). We will show that for any sequence $x^1, x^2, \ldots$ from X ,

$$(26) \qquad P_0 \left\{ \lim_{n\to\infty} \frac{\sup_{(V,\theta)\in Y} \Pi_{i=1}^n f(x^i, V, \theta)}{\Pi_{i=1}^n f(x^i, V^0, \theta^0)} = 0 \right\} = 1 .$$

It is clear that

$Y \subset \cup_{(V,\theta) \in Y} S(V, \theta, \varepsilon(V,\theta)) = \cup \{S(V, \theta, \varepsilon(V,\theta)) | (V,\theta) \in Y\}$
where $S(V, \theta, \varepsilon(V,\theta))$ denotes the sphere in $(W \times \Theta)$ with center (V,θ) and radius $\varepsilon(V,\theta)$.

Since Y is a closed subset of $(W \times \Theta)$ and by hypothesis $(W \times \Theta)$ is compact, Y is a compact set. Hence, there exists a finite sequence $\{(V_1,\theta_1),(V_2,\theta_2), \ldots, (V_h,\theta_h)\}$ in Y , and numbers $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_h$ such that $\varepsilon_k = \varepsilon(V_k,\theta_k)$ $(k=1,..,h)$ and

$$(27) \qquad Y \subseteq \cup_{k=1}^h S(V_k, \theta_k, \varepsilon_k) .$$

From (27) and (18) it follows that for all n and all $x^1, \ldots, x^n$

$$\sup_{(V,\theta)\in Y} \Pi_{i=1}^n f(x^i, V, \theta)$$

$$\leq \Sigma_{k=1}^h \sup_{(V,\theta) \in S(V_k,\theta_k,\varepsilon_k)} \Pi_{i=1}^n f(x^i, V, \theta)$$

$$\leq \Sigma_{k=1}^{h} \ \Pi_{i=1}^{n} \ f'(x^i, V_k, \theta_k, \varepsilon_k) \quad .$$

Hence,

$$(28) \qquad \frac{\sup_{(V,\theta) \in Y} \ \Pi_{i=1}^{n} \ f(x^i, V, \theta)}{\Pi_{i=1}^{n} \ f(x^i, V^0, \theta^0)}$$

$$\leq \frac{\Sigma_{k=1}^{h} \ \Pi_{i=1}^{n} \ f'(x^i, V_k, \theta_k, \varepsilon_k)}{\Pi_{i=1}^{n} \ f(x^i, V^0, \theta^0)}$$

Since by Kolmogorov's Strong Law of Large Numbers, and (25) for each $k = 1, \ldots, h$

$$(29) \quad P_0(\lim_{n\to\infty} \Sigma_{i=1}^{n}[\log f'(x^i, V_k, \theta_k, \varepsilon_k) - \log f(x^i, V^0, \theta^0)] = -\infty) = 1,$$

it follows that

$$(30) \qquad P_0 \left\{ \lim_{n\to\infty} \frac{\Pi_{i=1}^{n} \ f'(x^i, V_k, \theta_k, \varepsilon_k)}{\Pi_{i=1}^{n} \ f(x^i, V^0, \theta^0)} = 0 \right\} = 1 ,$$

for $k = 1, \ldots, h$.

By (30) and (28), (26) is proved.

Theorem 2 in Wald (1949) and (26) imply that

$$P_0 \ (\lim_{n\to\infty} r[(V_n, \theta_n), (V^0, \theta^0)] = 0) = 1 \ .$$

Q.E.D.

APPENDIX C

In this Appendix we present the proof of Lemma 1, stated in Section 4. We also state Lemmas 7-9, whose proofs are omitted because they follow closely the arguments of the proof of Lemma 1.

PROOF OF LEMMA 1: Let $D = \{0, z_1, \ldots, z_M\} \subset \text{int}(Z)$ be given. We first show that if $V \in W$ then $\underset{\sim}{V} = (V_1, \ldots, V_M) \equiv (V(z_1), \ldots, V(z_M))$ satisfies (7) and (8) for some $\underset{\sim}{T} = (T_0, T_1, \ldots, T_M) \in R^{K\times(M+1)}$.

Let $T^j = T(z_j)$ for $j = 0, 1, 2, \ldots, M$. It is clear from the definition of W that $T^j \leq B$ $(j=0,1,..,M)$ and that $\underset{\sim}{T} = (T^0, T^1, \ldots, T^M)$ and $\underset{\sim}{V} = (V(z_1), \ldots, V(z_M))$ satisfy (8). To see that $T^j \geq 0$ $(j=0,1,..,M)$ note that if w is such that $w \geq z_j$, the monotonicity and concavity of V imply that $0 \leq V(w) - V(z) \leq T(z_j)(w - z_j)$. But, since w can be chosen to be strictly greater than z_j in only one coordinate and equal to z_j in the other coordinates, it follows that $T_j = T(z_j) \geq 0$. So $\underset{\sim}{V}$ and $\underset{\sim}{T}$ satisfy (7) and (8).

We now show the converse, i.e., we show that if there exist $V_0=0, V_1, \ldots, V_M$ and $T^0, T^1, \ldots, T^M$ satisfying (7) and (8), there exists a function $V : Z \to R$ in W with $V(z_j) = V_j$ $(j=0,1,..,M)$.

Define the function $V : Z \to R$ by:

$$V(v) = \min\{V_j + T^j(v - z_j) \mid j = 0, 1, \ldots, M\} \tag{31}$$

for all $v \in Z$.

Then for all $j = 0, 1, \ldots, M$, $V(z_j) = V_j$. Since, suppose that $V(z_j) \neq V_j$ for some $j \in \{0, 1, \ldots, M\}$, then by (31) there exists $k \neq j$ such that $V(z_j) = V_k + T^k (z_j - z_k) < V$, contradicting one of the inequalities in (8).

It remains to show that $V \in W$. Since $V(z_j) = V_j$, $V(0)=0$. Moreover, by (31) and (7), V is monotonic and concave.

To show that V satisfies (ii), take any v and w in Z . By (31) there exists j in $\{0, 1, \ldots, M\}$ such that $V(v) = V_j + T^j(v - z_j)$ and $V(w) \leq V_j + T^j(v - z_j)$. Hence $V(w) - V(v) \leq T^j (w-v)$, and then V satisfies (ii) with $T(v) = T^j$.

So, V as defined in (31) is a monotonic and concave function satisfying (i), (ii), and $V(z_j) = V_j$ $(j = 0, 1, \ldots, M)$.

Q.E.D.

The following Lemmas present the relevant constraint sets for sets of functions $V : Z \to R$ that are monotonic, concave, with uniformly bounded subgradients, and that satisfy, in addition, one of the following sets of conditions: (a) homogeneity of degree one, (b) weak separability, and (c) additive separability.

LEMMA 7 (Homogeneity of degree one :) *Let W be the set of all concave and monotonic functions* $V : Z \to R$ *satisfying (i)-(ii) of Lemma 1, and (iii) for all* v *in* Z *and all* $\beta \in R_+$ *for which* $\beta v \in Z$, $V(\beta v) = \beta V(v)$.

Let $D = \{z_1, \ldots, z_M\}$ *be a finite subset of the interior of* Z . *Then the set of all vectors* $\underset{\sim}{V} = (V_1, \ldots, V_M)$ *for which there exists a function* V *in* W *with* $V_i = V(z_i)$ $(i = 1, \ldots, M)$ *is the set of all vectors* $\underset{\sim}{V} = (V_1, \ldots, V_M)$ *for which there exists a vector* $\underset{\sim}{T} = (T^0, T^1, \ldots, T^M) \in R^{K\times(M+1)}$ $(T^m = (T^m_1, \ldots, T^m_K) \in R^K,\ m = 0, 1, \ldots, M)$ *such that*

$$(32) \qquad 0 \leq T^j \leq B ,$$

and

$$(33) \qquad V^j \leq T^i z_i ,$$

for $i, j = 0, 1, \ldots, M$ *and for* $z_0 = 0$ *and* $V_0 = 0$.

Given any V_j , T^j $(j = 0, 1, \ldots, M)$ satisfying (32) and (33) with $V_0 = 0$, define the function $V : Z \to R$ by

$$(34) \qquad V(v) = \min \{ T^j v | j = 0, 1, \ldots, M \}.$$

for all $v \in Z$. The function V so defined is a function in W that interpolates between the values V_j and T^j .

LEMMA 8 (Weak separability :) *Let* T *be an integer* $(1 \leq T \leq K)$; *and let* $B^1 \in R^T_+$ *and* $B^2 \in R^{K-T}_+$ *be given. Define the sets* $Z^1 = Z \cap R^T_+$, $Z^2 = Z \cap R^{K-T}_+$ *and* $E = \{t \in R_+ | t \leq \max \{B^2 \|v^2\| \mid v^2 \in Z^2\} \}$.

Let W *be the set of all functions* $V : Z^1 \times Z^2 \to R$ *for which there exist two concave and monotonic functions* $u : Z^2 \to R$ *and* $y : (Z^1 \times E) \to R$

such that for all $v=(v^1, v^2) \in Z^1 \times Z^2$

(iv) $V(v) = V(v^1, v^2) = y(v^1, u(v^2))$,

(v) $u(0)=0$, $y(0)=0$, *and*

(vi) *for each* $v = (v^1, v^2)$ *in* $Z^1 \times Z^2$ *and* $t \in E$, *there exist subgradients* $T^u(v^2) \in R^{K-T}$ *and* $T^y(v^1, t) \in R^{T+1}$ *such that*

(vi.i) $T^u(v^2) \leq B^2$, *and* (vi.ii) $T^y(v^1, t) \leq B^1$.

Let $D = \{z_1, \ldots, z_M\}$ *be a finite subset of the interior of* Z. *Then the set of all vectors* $\underset{\sim}{V} = (V_1, \ldots, V_M)$ *for which there exists a function* V *in* W *with* $V_i = V(z_i)$ $(i = 1, \ldots, M)$ *is the set of all vectors* $\underset{\sim}{V} = (V_1, \ldots, V_M) = (y_1, \ldots, y_M)$ *for which there exist vectors* $(u_1, \ldots, u_M)$, $(T^{y0}, T^{y1}, \ldots, T^{yM})$ *and* $(T^{u0}, T^{u1}, \ldots, T^{uM})$ *such that*

$$0 \leq T^{uj} \leq B^2 , \tag{34}$$

$$0 \leq T^{yj} \leq B^1 , \tag{35}$$

$$u_i \leq u_j + T^{uj}(z_i^2 - z_j^2) , \text{ and} \tag{36}$$

$$y_i \leq y_j + T^{yj}((z_j^1, u_i) - (z_j^1, u_j)) , \tag{37}$$

for $i, j = 0, 1, \ldots, M$ *and for* $z_0 = (z_0^1, z_0^2) = (0,0)$, $y_0 = 0$, *and* $u_0 = 0$.

Given y_j, u_j, T^{yj} and T^{uj}, $j = 0, 1, \ldots, M$ satisfying (34)-(37), we can define the functions $u : Z^2 \to R$ and $T^u : Z^2 \to R^{K-T}$ by

$$u(v^2) = \min\{u_j + T^{uj}(v^2 - z_j^2) \mid j = 0, 1, \ldots, M\} , \tag{38}$$

and

$$T^u(v^2) = T^{uk} , \tag{39}$$

for all v^2 in Z^2, where k in (39) is the first h for which $u(v^2) = u_h + T^{uh}(v^2 - z_h^2)$; and we can define the functions $y : Z^1 \times E \to R$ and $T^y : Z^1 \times E \to R^{T+1}$ by

$$(40) \qquad y(v^1, t) = \min(y_j + T^{yj}((v^2, t) - (z_j^2, u_j)) | j = 0, 1, \ldots, M) ,$$

and

$$(41) \qquad T^y(v^2, t) = T^{yk} ,$$

for all t in E and v^1 in Z^1, where k in (41) is the first h for which $y(v^2, t) = y_h + T^{yh}((v^2, t) - (z_h^2, u))$.

Defining $V : Z^1 \times Z^2 \to R$ by

$$(42) \qquad V(v^1, v^2) = y(v^1, u(v^2))$$

for all $v = (v^1, v^2)$ in $Z^1 \times Z^2$, we obtain a function that belongs to W, and interpolates between the obtained values for V and subdifferentials of V at $z_0, z_1, \ldots, z_M$.

LEMMA 9 (Additive separability :) *Let T be an integer* $(1 \le T \le K)$; *and let* $B^1 \in R_+^T$ *and* $B^2 \in R_+^{K-T}$ *be given. Define, as in Lemma 8, the sets* $Z^1 = Z \cap R_+^T$ *and* $Z^2 = Z \cap R_+^{K-T}$. *Let W be the set of all functions* $V : Z^1 \times Z^2 \to R$ *for which there exist two monotonic and concave functions* $y^1 : Z^1 \to R$ *and* $y^2 : Z^2 \to R$, *such that for all* $v = (v^1, v^2) \in Z^1 \times Z^2$

(vii) $V(v) = V(v^1, v^2) = y^1(v^1) + y^2(v^2)$,

(viii) $y^1(0)=0$, $y^2(0)=0$, and

(ix) for each $v = (v^1, v^2)$ in $Z^1 \times Z^2$, there exist subgradients $T^{y1}(v^1) \in R^T$ of y^1, and $T^{y2}(v^2) \in R^{K-T}$ of y^2 such that $T^{y1}(v^1) \le B^1$, and $T^{y2}(v^2) \le B^2$.

Let $D = \{z_1, \ldots, z_M\}$ be a finite subset of the interior of Z . Then the set of all vectors $\underset{\sim}{V} = (V_1, \ldots, V_M)$ for which there exist a function V in W with $V_i = V(z_i)$ $(i = 1, \ldots, M)$ is the set of all vectors $\underset{\sim}{V} = (V_1, \ldots, V_M)$ for which there exist vectors $(y_1^1, \ldots, y_M^1)$, $(y_1^2, \ldots, y_M^2)$, $(T^{y1,0}, T^{y1,1}, \ldots, T^{y1,M})$ and $(T^{y2,0}, T^{y2,1}, \ldots, T^{y2,M})$ such that

$$0 \le T^{y1,j} \le B^1 , \tag{43}$$

$$0 \le T^{y2,j} \le B^2 , \tag{44}$$

$$y_i^1 \le y_j^1 + T^{y1,j}(z_i^1 - z_j^1) , \tag{45}$$

$$y_i^2 \le y_j^2 + T^{y2,j}(z_i^2 - z_j^2) , \tag{46}$$

and

$$V_j = y_j^1 + y_j^2 \tag{47}$$

for all $i, j = 0, 1, \ldots, M$, and for $z_0 = (z_0^1, z_0^2) = (0,0)$, $y_0^1 = 0$, and $y_0^2 = 0$.

Given V_j , y_j^1 , y_j^2 , $T^{y1,j}$, $T^{y2,j}$, $j = 0, 1, \ldots, M$ satisfying (43)-(47) we can define the functions $y^1 : Z^1 \to R$, $T^{y1} : Z^1 \to R^T$,

$y^2 : Z^2 \to R$ and $T^{y2} : Z^2 \to R^{K-T}$, by

(48) $$y^1(v^1) = \min\{y^1_j + T^{y1,j}(v^1 - z^1_j) \mid j = 0, 1, \ldots, M\} ,$$

(49) $$T^{y1}(v^1) = T^{y1,k} ,$$

where k is the first h for which $y^1(v^1) = y^1_h + T^{y1,h}(v^1 - z^1_h)$,

(50) $$y^2(v^2) = \min\{y^2_j + T^{y2,j}(v^2 - z^2_j) \mid j = 0, 1, \ldots, M\} ,$$

and

(51) $$T^{y2}(v^2) = T^{y2,k} ,$$

where k is the first h for which $y^2(v^2) = y^2_h + T^{y2,h}(v^2 - z^2_h)$.

Then, by defining $V : Z^1 \times Z^2 \to R$ by

$$V(v^1, v^2) = y^1(v^1) + y^2(v^2) ,$$

for all (v^1, v^2) in $Z^1 \times Z^2$, we obtain a function that belongs to W , interpolates between the obtained values for V_j , and its subgradient interpolates between the obtained values for $(T^{y1,j}, T^{y2,j})$.

REFERENCES

AFRIAT, S. (1967a), "The Construction of a Utility Function from Demand Data," *International Economic Review*, Vol. 8.

_____ (1967b): "The Construction of Separable Utility Functions from Expenditure Data," mimeo, Purdue University.

_____ (1972): "The Theory of International Comparison of Real Income and Prices," in J. D. Daly (ed.), *International Comparisons of Prices and Output*. New York: National Bureau of Economic Research.

_____ (1973): "On a System of Inequalities on Demand Analysis," *International Economic Review*, Vol. 14.

_____ (1981): "On the Constructability of Consistent Price Indices between Several Periods Simultaneously," in A. Deaton (ed.) *Essays in Applied Demand Analysis*. Cambridge University Press.

COSSLETT, S. R. (1983): "Distribution-free Maximum Likelihood Estimator of the Binary Choice Model," *Econometrica*, Vol. 51, No. 3, pp. 765-782.

DIEWERT, E. W. (1973): "Afriat and Revealed Preference Theory," *Review of Economic Studies*, 40.

EPSTEIN, L. G. and D. J. YATCHEW (1985): "Non-Parametric Hypothesis Testing Procedures and Applications to Demand Analysis," *Journal of Econometrics*, 30.

GALLANT, A. R. (1981): "On the Basis in Flexible Functional Forms and an Essentially Unbiased Form," *Journal of Econometrics*, No. 15, pp. 211-245.

_____ (1982): "Unbiased Determination of Production Technologies," *Journal of Econometrics*, 20.

GALLANT, A. R. and G. H. GOLUB (1984): "Imposing Curvature Restrictions on Flexible Functional Forms," *Journal of Econometrics*, 26, pp. 295-321.

HAN, A. K. (1985a): "Nonparametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator", Discussion Paper 1173, Harvard Institute of Economic Research, Harvard University.

_____ (1985b): "Large Sample Properties of Nonparametric Estimators in Generalized Regression Models", Discussion Paper 1199, Harvard Institute of Economic Research, Harvard University.

KIEFER, J. and J. WOLFOWITZ (1956): "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," *Annals of Mathematical Statistics*, Vol. 27, pp. 887-906.

MANSKI, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, Vol. 3, No. 3, pp. 205-228.

_____ (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator", *Journal of Econometrics*, Vol. 27, pp. 313-334.

MATZKIN, R. L. (1986): *Mathematical and Statistical Inferences from Demand Data*, Ph. D. Thesis, University of Minnesota.

McFADDEN, D. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), *Frontiers of Econometrics* New York: Academic Press, pp. 105-142.

_____ (1976): "Quantal Choice Analysis: A Survey," *Annals of Economics and Social Measurements*, 5, pp. 363-390.

_____ (1978): "Modelling the Choice of Residential Location," in A. Karlquist et al. (eds.), *Spatial Interaction Theory and Residential Location*, pp. 75-96. Amsterdam: North-Holland.

_____ (1981): "Econometric Models of Probabilistic Choice," in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data*. Cambridge: MIT PRess, pp. 198-272.

_____ (1984): "Econometric Analysis of Qualitative Response Models", in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, Vol. II, North Holland.

VARIAN, H. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, Vol. 50, No. 4.

_____ (1983): "Nonparametric Tests of Consumer Behavior", *Review of Economic Studies*, No. 50.

_____ (1985): "Non-Parametric Analysis of Optimizing Behavior with Measurement Error," *Journal of Econometrics*, 30.

WALD, Abraham (1949): "Note on the Consistency of the Maximum Likelihood Estimate," *Annals of Mathematical Statistics*, Vol. 20, pp. 595-601.

NOTES

1 An earlier version of this paper was part of my Ph.D Thesis, and was presented at the 1986 summer meetings of the Econometric Society under the title "Nonparametric Estimation of Utility Functions for Discrete Choice Models." I am indebted to Professors Lung-Fei Lee, Marcel K. Richter and Christopher Sims for their valuable advice. I wish to thank Philip Dybvig, Vassilis Hajivassiliou, and Alvin Klevorick for their comments.

2 If the set of alternatives that maximize $V^0(\ ) + e(\ )$ possesses h elements, j(i) is assumed to be drawn from this set of maximal elements with probability 1/h. Of course, if the e^i_j terms are distributed with a continuous distribution there will be, with probability 1, a unique maximizer j(i) in {1,..,J}.

3 If the distribution of $(e^i_1,..,e^i_J)$ is not continuous, we let $\{A_t | t=1,..,2^J\}$ be the set of all subsets of A, then

$$P^i_j(V^0,\theta^0)=\Sigma^{2^J}_{t=1}\beta(j,A_t)\ (1/\#(A_t))\ \text{Prob}(A_t=\{j\epsilon A | V^0(z^i_j)+e^i_j \geq V^0(z^i_k)+e^i_k,\ k\epsilon A\})$$

where $\beta(j,A_t)=1$ if $j\epsilon A_t$, and $\beta(j,A_t)=0$ otherwise; and where $\#(A_t)$ denotes the number of elements in A_t.

4 This is commonly the case when each individual consumer is observed only once.

5 The Gumbel distribution is defined by $P(\epsilon \leq \eta) = \exp(-\exp(-\eta))$.

6 Matzkin(1986) employed these computation methods to estimate monotonic, and monotonic and concave utility functions from simulated data on discrete choice behavior.

7 A function $V : Z \rightarrow R$ is said to be *monotonic* if for all $v, w \in Z$, $v \leq w$ implies $V(v) \leq V(w)$; $V : Z \rightarrow R$ is said to be *concave* if for all $v, w \in Z$ and all $\lambda \in [0,1]$, $V(\lambda v+(1-\lambda)w) \geq \lambda V(v)+(1-\lambda)V(w)$.

8 A vector $T(v) \in R^K$ is a *subgradient* of a function $V : Z \rightarrow R$ at $v \in Z$ if for all $w \in Z$, $V(w) - V(v) \leq T(v)(w-v)$.

9 The *subdifferential* of a concave function $V : Z \rightarrow R$ at $v \in Z$, is the set of all subgradients of V at v.