Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers

Cowles Foundation

10-1-1990

Estimation of Multinomial Models Using Weak Monotonicity Assumptions

Rosa L. Matzkin

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

Part of the Economics Commons

Recommended Citation

Matzkin, Rosa L., "Estimation of Multinomial Models Using Weak Monotonicity Assumptions" (1990). *Cowles Foundation Discussion Papers*. 1200.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/1200

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 957

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

ESTIMATION OF MULTINOMIAL MODELS USING WEAK MONOTONICITY ASSUMPTIONS

Rosa L. Matzkin

October 1990

ESTIMATION OF MULTINOMIAL MODELS USING WEAK MONOTONICITY ASSUMPTIONS

by

Rosa L. Matzkin*
Cowles Foundation
Yale University

ABSTRACT

This paper introduces a semiparametric method of estimating multinomial models that imposes extremely weak monotonicity assumptions about a function of observable characteristics. Previous methods have imposed stronger, typically parametric, conditions on this function. The only assumptions made in this paper about the function of characteristics are its monotonicity, upper-semicontinuity, and uniform boundedness. The method is applicable, among others, to polychotomous choice models. The estimation method is shown to be strongly consistent. A technique to calculate the estimator is provided.

1. INTRODUCTION

A semiparametric estimation method for multinomial models is developed that requires extremely weak monotonicity assumptions about a function, V, of characteristics of the outcomes. Weak assumptions lessen the possibility of specification errors. This is important because even small specification errors may produce large inconsistencies in estimated parameters (Abramazar and Schmidt (1982), Hausman and Wise (1978)).

In this paper, we require the nonparametric vector-valued function V to be only monotone increasing, upper-semicontinuous, and uniformly bounded; not even its continuity is assumed. The probability of each outcome is assumed to be a known parametric function of V . Probabilities with such structures arise frequently in polychotomous choice models when the distribution of the unobservable random subutilities is parametric.

The assumptions we make about the function V are much weaker than those made in Matzkin (1989), where a very particular case of multinomial models - polychotomous choice models - was studied. In Matzkin (1989) the function V was required to be monotone, concave, continuous, and to possess uniformly bounded values and subgradients. Moreover, strong restrictions were imposed on the dimensionality of V and on the probabilities of each outcome.

There are at least three reasons why it is important to study the estimation of the function V subject to the much weaker monotonicity restrictions considered in this paper. First, it is of theoretical interest to determine the extent to which the assumptions can be relaxed while still obtaining strong consistency results. Second, the computation of the monotonicity-restricted estimator presented in this paper involves solving a smaller optimization problem than the one required to compute the concavity-restricted estimator. Hence, the computation of the estimator presented here is cheaper. Third, when the function V is known to satisfy stronger restrictions than the ones required in this paper, one can employ the results from Monte Carlo experiments obtained for the monotonicity-restricted estimator to provide a lower bound on the performance of estimators that impose the stronger restrictions.

Our estimator is obtained by maximizing the likelihood function over a set of nonparametric functions and a set of parameter vectors. The estimator of V belongs to the former set while the estimator of the parameter, θ, belongs to the latter set. The maximum likelihood estimator of the nonparametric function V and the parameter vector θ is shown to be strongly consistent.

The estimator proposed in this paper is dual, in a sense, to the estimator proposed in Cosslett (1983). Both are obtained by maximizing a likelihood function over a set of nonparametric monotone functions and a set of parameter vectors. In Cosslett's model the probability function belongs to the set of nonparametric functions; here the function of exogenous variables belongs to this set.

There are, however, some more differences between Cosslett's model and ours. First, Cosslett's estimator has only been developed for the case of two alternatives; our estimator can be applied to models in which there is any finite number of alternatives. In particular, our estimator can be employed to predict choice probabilities when a new alternative becomes available. Second, Cosslett assumes that the probability function does not depend directly on the observable exogenous variables; our model does not require such a restriction. This is an important distinction because neglecting the dependence on r may severely bias the estimators (Hausman and Wise (1978)). Third, the computation of Cosslett's estimator requires the maximization of a discontinuous function; our estimator is calculated by maximizing a continuous, typically well behaved, function subject to a finite number of inequality constraints. Hence, standard methods of maximization can be employed. Fourth, in Cosslett's model, the function V is real-valued; in our model V is vector-valued. The probabilities of each outcome may depend on the values of any finite number of coordinate functions. This is important, for example, in choice models in which the value of each alternative depends on alternative specific constants.

Besides the estimators of Cosslett (1983) and Matzkin (1989) mentioned before, there are other semiparametric estimators that have been developed for either binomial or multinomial models. These include the maximum score estimator (Manski (1975,1985)), the quasi-maximum likelihood estimator of Klein and Spady (1988), and the generalized regression models of Stoker (1986), Han (1987), Ichimura (1988) and Powell, Stock, and Stoker (1990). These are distribution-free methods; they do not assume that the probability functions of the outcomes are known. They impose, however, parametric assumptions about the function V. Hence, these methods are likely to misspecify V.

Fully-nonparametric methods (Matzkin (1988,1990a,1990b)) avoid misspecifications of both V and the probability functions. These methods typically require, however, stronger conditions on V and on the probability functions than the ones required by the method presented in this paper.

The outline of the paper is as follows: The model and the estimator are defined in the next section. The identification of the nonparametric function V of the observable exogenous variables and the parameter θ of the probability functions is analyzed in Section 3; the consistency of the estimator is studied in Section 4. Section 5 presents and analyzes a method of computing the estimator; and Section 6 summarizes the main results and conclusions. The proofs of the theorems are in the Appendix.

2. THE MODEL

In this model, there is a finite set A of J alternative outcomes. The probability of observing each outcome depends on a vector of observable characteristics, r, the value that a vector-valued function, V^*, attains at the vector r, and a finite-dimensional parameter vector θ^*. The probability of observing outcome $j \in A$ when the vector of observable characteristics is r will be denoted by $P(j|r;V^*,\theta^*)$.

The value of each coordinate function V_t^* of V^* depends on the value of subvectors, s and z_t, of r, and is independent of the other subvectors of r. The vector r is $(s,z_1,\ldots,z_T)$, where T is the number of coordinate functions of V. We denote by S and $Z=\Pi_{t=1}^T Z_t$, respectively, the sets to which s and $z=(z_1,\ldots,z_T)$ belong.

The vector $r = (s,z)$ possesses a probability density g determined by a probability measure G, whose support is $S \times Z$.

The probability functions described above are encountered, for example, in *polychotomous choice models* (McFadden (1973,1981)). In these models, an outcome corresponds to the selection of one of J alternatives by a consumer. The vector s denotes socioeconomic characteristics of the consumer and each vector z_j denotes the vector of the observable attributes of alternative j. The consumer is assumed to select the alternative that maximizes the value of his utility function. The consumer's utility for each alternative j is the sum of $V_j^*(s,z)$ and an unobservable random term ε_j. The vector $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_J)$ is assumed to possess, for each (s,z), a density $q(\varepsilon;s,z,\theta^*)$, which is known up to the vector θ^*. Hence, in these polychotomous choice models,

$$P(t|s,z;V^*,\theta^*)$$

$$= Prob\ (\ V_t^*(s,z) + \varepsilon_t > V_k^*(s,z) + \varepsilon_k\ ;\ k\neq t,\ k=1,\ldots,J\ )$$

(1)

$$= \int_{\varepsilon_t=-\infty}^{\infty} \int_{\varepsilon_1=-\infty}^{V_t^*-V_1^*+\varepsilon_t} \cdots \int_{\varepsilon_J=-\infty}^{V_t^*-V_J^*+\varepsilon_t} q(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_J;s,z,\theta^*)\ d\varepsilon_1 \cdots d\varepsilon_J\ ,$$

where $V_j^* = V_j^*(s,z)\quad (j=1,\ldots,J)$.

In this paper we are concerned with estimating the function V^* and the parameter vector θ^* in general multinomial models. Instead of assuming that V^* is known up to a finite dimensional parameter vector, as most econometric methods do, we impose only very weak assumptions about V^*. We require V^* to be monotone increasing, upper-semicontinuous, and uniformly bounded[1,2]. To insure the identification of the function V^*, additional conditions are required. These conditions require that for some value, $\bar{z}$, of z and all values $s \in S$ the value of $V^*(s,\bar{z})$ is known and constant.

Our estimator is developed for the case in which n independent observations on the vector of characteristics (s,z) and the corresponding outcome are available. In this case, the conditional log-likelihood function is

$$(2) \qquad \sum_{i=1}^{n} \sum_{j=1}^{J} d_j^i \log P(j|s,z;V,\theta)$$

where for each i, $d_j^i=1$ if outcome j occurs in observation i and $d_j^i=0$ otherwise.

To define an appropriate set, W, of nonparametric functions to which V^* belongs, we let $\bar{z} \in Z$ and $\alpha=(\alpha_1,...,\alpha_J) \in R^K$ be known, and we let $L = (L_1,...,L_J)$ and $U = (U_1,...,U_J)$ denote, respectively, the lower and upper bound of V^*. We then define the set W to be the set of monotone increasing, upper-semicontinuous functions $V: S \times Z \to R^J$ such that V is bounded below by L, bounded above by U, and $V(s,\bar{z}) = \alpha$ for all $s \in S$. We also let Θ be a compact set which contains θ^*. We then define our semiparametric estimator to be any pair (V_n,θ_n) that maximizes (2) over the set $(W \times \Theta)$.

3. IDENTIFICATION

The pair (V^*, θ^*) will be identified within $(W \times \Theta)$ if (V^*, θ^*) belongs to $(W \times \Theta)$ and if any other (V, θ) that belongs to $(W \times \Theta)$ induces conditional probabilities that are different from the true conditional probabilities. Following Manski (1988), we state this formally by saying that (V^*, θ^*) is <u>identified</u> within $(W \times \Theta)$ if for all $(V, \theta) \in (W \times \Theta)$ such that $(V, \theta) \neq (V^*, \theta^*)$ there exists $j \in \{1, \ldots, J\}$ and $Y(V, \theta) \subset S \times Z$ with $G(Y(V, \theta)) > 0$ such that for all $(s, z) \in Y(V, \theta)$, $P(j|s, z; V, \theta) \neq P(j|s, z; V^*, \theta^*)$.[3]

We will show that (V^*, θ^*) is identified in $(W \times \Theta)$ whenever the following assumptions are satisfied.

ASSUMPTION Θ.1: $\theta^* \in \Theta \subset R^L$.

ASSUMPTION W.1: *W is a set of monotone increasing, upper-semicontinuous functions* $V: S \times Z \rightarrow R^T$, *such that for all* $(s, z) \in S \times Z$ $L \leq V(s, z) \leq U$.

ASSUMPTION W.2: $V^* \in W$.

ASSUMPTION W.3: $\forall V \in W$, $\forall (s, z) \in S \times Z$, $\forall t$, *and* $\forall k \neq t$ $V_t(s, z)$ *does not depend on* z_k.

ASSUMPTION W.4: *There exists* $(\alpha_1, \ldots, \alpha_T) \in R^T$ *and* $\bar{z} \in Z$ *such that* $\forall$ $V \in W$ *and* $\forall$ $s \in S$, $V(s, \bar{z}) = (\alpha_1, \ldots, \alpha_T)$.

ASSUMPTION G.1: $\forall (s,z) \in S \times Z$ and $\forall\ \delta > 0$ $G[N((s,z),\delta) \cap U(s,z)] > 0$, where $N(s,z) = \{(s',z') \mid \|(s',z')-(s,z)\| < \delta\}$ and $U(s,z) = \{(s'z') \mid (s',z') \geq (s,z)\}$.

ASSUMPTION P.1: *For all* $j \in A$ *there exists a function* $\bar{P}_j : S \times Z \times [L,U] \times \Theta \to R$ *such that* $\forall (s,z) \in S \times Z$, *and* $\forall V \in W$ $P(j \mid s,z;V,\theta) = \bar{P}_j(s,z;V_1(s,z),\ldots,V_T(s,z),\theta)$.

ASSUMPTION P.2: $\forall j \in A$, $\forall (s,z) \in S \times Z$, $\forall V \in W$, *and* $\forall \theta \in \Theta$, $\bar{P}_j(s,z;V_1,\ldots,V_T,\theta)$ *is continuous at* $(s,z,V_1,\ldots,V_T,\theta)$.

ASSUMPTION P.3: $\forall t \in \{1,\ldots,T\}$ *there exists* $j \in A$ *such that* $\forall (s,z) \in S \times Z$, $\forall V \in W$, *and* $\forall \theta \in \Theta$, *the value of* $P(j \mid s,z;V,\theta)$ *is either strictly increasing in the value of* $V_t(s,z)$ *or strictly decreasing in the value of* $V_t(s,z)$.

ASSUMPTION P.4: $\forall j \in A$, $\forall (s,z) \in S \times Z$, $\forall V \in W$, *and* $\forall \theta \in \Theta$, $P(j \mid s,z;V,\theta) > 0$.

ASSUMPTION P.5: $\forall\ \theta \in \Theta$ *such that* $\theta \neq \theta^*$ *there exists* $(s,z) \in S \times Z$ *and* $j \in A$ *such that* $P(j \mid s,z;V^*,\theta^*) \neq P(j \mid s,z;V^*,\theta)$.

Assumption W.1 states that each coordinate function V_t of any function V in W is a monotone increasing and upper-semicontinuous function that is uniformly bounded by L_j and U_j. The number, T, of coordinate functions is the same for all functions in W, and it does not necessarily depend on the number, J, of possible outcomes. By Assumption W.3, each coordinate function, V_t, possesses as an argument a vector z_t, which does not influence the values of the other coordinate functions. The identification of each coordinate of V^* depends on this condition, as well as on Assumption W.4. Since, these two assumptions, together with Assumption W.2, guarantee that for any function V in W that is different from V^*, there exists a vector (s,z) and a coordinate t such that $V_t(s,z)$ is different from $V_t^*(s,z)$ while for all other coordinates k, $V_k(s,z)$ equals $V_k^*(s,z)$. The strict monotonicity with respect to V_k of the probability of some outcome (Assumption P.3) implies then, that the probabilities generated by (V,θ^*) at (s,z) are different from the probabilities generated by (V^*,θ^*) at (s,z).[4]

Assumption P.1 states that the probability function $P(j|s,z;V,\theta)$ of each outcome j is a function of the observable vectors (s,z), the value of V at (s,z), and θ. In particular, this assumption implies that values of V at points other than (s,z) do not influence the value of $P(j|s,z;V,\theta)$. Assumption P.2 states that the value of $P(j|s,z;V,\theta)$ depends continuously on s,z,θ, and the value of $V(s,z)$. We employ Assumptions P.1 and P.2 to show the existence of a neighborhood in $S\times Z$ at which the probabilities generated by V^* and V differ when V and V^* differ. Assumption G.1 guarantees that the probability of this neighborhood is positive. To insure the identification of θ^*, we make Assumption P.5.

Assumption P.1 is always satisfied when the probabilities are generated by (1), and Assumptions P.2 and P.3 follow, when (1) is satisfied, if for any $(s,z)\in S\times Z$, the set $\{ y\in R^T \mid L \le y \le U \}$ is strictly included in the support of $q(\varepsilon;s,z,\theta^*)$.

Our main result in this section is the following theorem:

THEOREM 1: *Suppose that Assumptions* $\Theta.1$, *W.1-W.4, G.1, and P.1-P.5 are satisfied. Then* (V^*,θ^*) *is identified within* $(W\times\Theta)$.

4. CONSISTENCY

In this section we show the strong consistency of our estimator. This result is obtained with respect to the metric $d:(W\times\Theta) \times (W\times\Theta) \rightarrow R$ defined by

$$d[(V,\theta),(V',\theta)] = m(V,V') + \|\theta-\theta'\|,$$

where for all $V,V'\in W$

$$(3) \qquad m(V,V') = \int \| V(s,z) - V'(s,z) \| \ \exp(\|z\|) \ dz \ ;$$

the integration is with respect to the Lebesgue measure[5]. Convergence in W with respect to m is equivalent to pointwise convergence at all points of continuity of the limit function (see Appendix, proofs of Lemmas 2 and 3).

In the Appendix we show that the set of functions W is compact with respect to m (Lemma 1).

To prove the strong consistency result, we use the assumptions stated in the previous section together with the following assumptions.

ASSUMPTION Θ.2: *Θ is compact with respect to* $\|\cdot\|$

ASSUMPTION G.2: *The support of G is S×Z .*

ASSUMPTION G.3: *G is absolutely continuous.*

ASSUMPTION G.4: *The probability density g is uniformly bounded.*

The compactness of $W \times \Theta$ will allow us to employ an adaptation of the result of Wald (1949) about the consistency of maximum likelihood estimators. If the support of G is strictly included in the domain of the functions in W, it will be impossible to estimate the values of V^* at points outside the support of G. Assumption G.2 insures that this is not the case. To guarantee that the set of points (s,z) at which the probabilities are discontinuous will possess zero probability measure, we require G to be absolutely continuous (Assumption G.3). Assumptions W.1 and P.2 insure that the set of points of discontinuity of the probabilities possess zero Lebesgue measure. These two latter assumptions also guarantee, together with Assumption P.1 and the compactness of W, that the probabilities are measurable in (s,z). In particular, Assumption P.1 allow us to find, for each (s,z) in $S \times Z$ and V in W, a function V' in W that is continuous at (s,z) and attains the same value as V does at (s,z). This guarantees that any sequence of a dense set converging to V' with respect to m will converge pointwise at (s,z) to V. Finding such a sequence is a critical step in proving the measurability of the probability functions.

Our consistency result is stated in the next theorem, which is proved in the Appendix.

THEOREM 2: *Suppose that Assumptions Θ.1-Θ.2, W.1-W.4, G.1-G.4, and P.1-P.5 are satisfied. Then,*

$$Pr\left\{ \lim_{n\to\infty} d\,[(V_n, \theta_n),(V^*,\theta^*)] = 0 \right\} = 1 .$$

We next present an example of a polychotomous choice model satisfying all the assumptions made above.

EXAMPLE: Suppose that T=J=3. The vector (s,z) possesses a normal density function. W is the set of all monotone increasing, upper-semicontinuous functions $V: R^3 \to R^3$, such that for all $z=(z_1,z_2,z_3) \in R^3$ the value of V_i depends solely on z_i, lies in $[-1,1]$, and satisfies $V(0) = 0$. For some $\delta < 1$, $\Theta=[0,\delta]$, and $0 < \theta^* < \delta$. And the conditional probabilities of outcomes 1, 2, and 3 given z, at any $V \in W$ and $\theta \in \Theta$, are:

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

$$P(2|z;\ V,\theta) = \frac{\exp(V_2)^{1/(1-\theta)} \quad [(\exp(V_2)^{1/(1-\theta)} + \exp(V_3)^{1/(1-\theta)})^{-\theta}]}{[\exp(V_1) + [\exp(V_2)^{1/(1-\theta)} + \exp(V_3)^{1/(1-\theta)}]^{(1-\theta)}]};\ \text{and}$$

$$P(3|z;\ V,\theta) = \frac{\exp(V_3)^{1/(1-\theta)} \quad [(\exp(V_2)^{1/(1-\theta)} + \exp(V_3)^{1/(1-\theta)}]^{-\theta}]}{[\exp(V_1) + [\exp(V_2)^{1/(1-\theta)} + \exp(V_3)^{1/(1-\theta)}]^{(1-\theta)}]};$$

where $V_j=V_j(z)$ for j=1,2,3.[6]

It is then easy to verify that this example satisfies Assumptions Θ.1-Θ.2, W.1-W.4, G.1-G.4, and P.1, P.2, and P.4. Assumption P.3 is satisfied because when z = 0, the partial derivative of $P(1|z;\ V,\theta)$ with respect to

V_1 is strictly positive and the partial derivatives of $P(1|z; V,\theta)$ with respect to V_2 and V_3 are strictly negative. Finally, since when $z = 0$ and $\theta = \theta^*$, the partial derivative of $P(1|z; V,\theta)$ with respect to θ is strictly negative[7], Assumption P.5 is satisfied.

5. COMPUTATION

To calculate the estimator, we transform the maximization over the set $W\times\Theta$ into a constrained maximization over a finite dimensional euclidean space[8]. The variables over which the maximization takes place are the parameter vector θ and the vectors $V(s^1,z^1),\ldots,V(s^n,z^n)$. The values of $V(s^1,z^1),\ldots,V(s^n,z^n)$ need to satisfy properties that characterize the set of values that any function V in W may attain at each of the observed pairs (s^i,z^i) $(i=1,\ldots,n)$.

To guarantee that the values of $V_t(s^1,z^1),\ldots,V_t(s^n,z^n)$ belong to a monotone increasing function, we require that these values be monotone increasing with respect to the set $\{ (s^1,z^1_t),\ldots,(s^n,z^n_t) \}$. That is, we require that whenever all the coordinates of (s^q,z^q_t) are smaller than the corresponding coordinates of (s^r,z^r_t), the value of $V_t(s^q,z^q_t)$ be smaller than the value of $V_t(s^r,z^r_t)$.

To insure that the values of $V_t(s^1,z^1),\ldots,V_t(s^n,z^n)$ belong to a function that attains the value α_t at any vector $(s,\bar{z}_t)$, we first find vectors s^* and s_* that are, respectively, a lower and upper bound of the set $\{s^1,\ldots,s^n\}$. Next, we add the vectors $(s^*,\bar{z}_t)$ and $(s_*,\bar{z}_t)$ and the variables $V_t(s^*,\bar{z}_t)=\alpha_t$ and $V_t(s_*,\bar{z}_t)=\alpha_t$, respectively, to the sets

$\{(s^1,z_t^1),\ldots,(s^n,z_t^n)\}$ and $\{V_t(s^1,z^1),\ldots,V_t(s^n,z^n)\}$. Finally, we require that the values of $\{V_t(s^1,z^1),\ldots,V_t(s^n,z^n),V_t(s^*,\bar{z}_t),V_t(s_*,\bar{z}_t)\}$ be monotone increasing with respect to the set $\{(s^1,z_t^1),\ldots,(s^n,z_t^n),(s^*,\bar{z}_t),(s_*,\bar{z}_t)\}$. Doing this guarantees that any monotone function that interpolates between the so restricted values of $\{V_t(s^1,z^1),\ldots,V_t(s^n,z^n),V_t(s^*,\bar{z}_t),V_t(s_*,\bar{z}_t)\}$ will attain the value α_t at any vector $(s,\bar{z}_t)$ for which $s_* \le s \le s^*$.

To guarantee that the values of $V_t(s^1,z^1),\ldots,V_t(s^n,z^n)$ belong to a function that is uniformly bounded by L_t and U_t, we augment the sets $\{(s^1,z_t^1),\ldots,(s^n,z_t^n)\}$ and $\{V_t(s^1,z^1),\ldots,V_t(s^n,z^n)\}$ once more. This time we add vectors (s^*,z_t^*) and (s_*,z_{*t}) to the former set and variables $V_t(s^*,z_t^*)$ and $V_t(s_*,z_{*t})$ to the latter set. The vectors z_t^* and z_{*t} are, respectively, an upper and lower bound of the set $\{ z_t^1,\ldots,z_t^n,\bar{z}_t,\bar{z}_t \}$. The values of $V_t(s^*,z_t^*)$ and $V_t(s_*,z_{*t})$ are restricted to be L_t and U_t respectively. By requiring that the values of $\{V_t(s^1,z^1),\ldots,V_t(s^n,z^n),V_t(s^*,\bar{z}_t),V_t(s_*,\bar{z}_t),V_t(s^*,z_t^*), V_t(s_*,z_{*t}) \}$ be monotone increasing with respect to $\{(s^1,z_t^1),\ldots,(s^n,z_t^n),(s^*,\bar{z}_t),(s_*,\bar{z}_t),(s^*,z_t^*),(s_*,z_{*t})\}$, we guarantee that the values of any monotone interpolation V_t will lie between L_t and U_t, for any vector (s,z) lying between (s^*,z_t^*) and (s_*,z_{*t}).

To state these constraints formally, we next define for each t the augmented set T_t. We let s^* and s_* be such that for all i=1,...,n

$$s_* \le s^i \le s^*$$

and for all t we let z_t^* and z_{*t} be such that for all i

$$z_t^* \le z_t^i \le z_{*t}^i.$$

We then define

$$T_t = \{ (s^1,z_t^1),\ldots,(s^n,z_t^n),(s^*,\bar{z}_t),(s_*,\bar{z}_t),(s^*,z_t^*),(s_*,z_{*t}) \}.$$

The sets T_t are the sets of points on which we will impose the monotonicity restrictions.

To avoid imposing redundant monotonicity constraints, however, we further define some subsets of the T_t sets. For each t and each element y of T_t, we define the set $F_j(y)$ of "immediate followers" of y by

$F_j(y) = \{ w \in T_j \mid y \le w$ and for no $z \in T_j$ such that $z \neq w$ and $y \le z,\ z \le w \}$.

Employing this notation, we next describe in Theorem 3 the procedure to calculate the estimator.

THEOREM 3: (V,θ) maximizes (2) over the set $(W \times \Theta)$ if and only if the vector $(V_1(s^1,z^1),\ldots,V_T(s^1,z^1);\ldots;V_1(s^n,z^n),\ldots,V_T(s^n,z^n);\theta)$ is a solution to the maximization of

$$(4) \qquad \sum_{i=1}^{n} \sum_{j=1}^{J} d_j^i \log \bar{P}_j(s^i,z^i;V_1^i,\ldots,V_T^i,\theta) \ .$$

subject to the constraints

$$(5) \quad V_t^i \le V_t^r \qquad \forall (s^r,z_1^r) \in F_1(s^i,z_1^i) \qquad i=1,\ldots,n,n+1,n+2;\ t=1,\ldots,T,$$

$$(6) \quad V_t^i \le U_t \qquad \text{if } (s^*,z_t^*) \in F_1(s^i,z_1^i) \qquad i=1,\ldots,n,n+1,n+2;\ t=1,\ldots,T,$$

$$(7) \quad L_t \le V_t^i \qquad \forall (s^i,z_1^i) \in F_1(s_*,z_{*t}) \qquad i=1,\ldots,n,n+1,n+2;\ t=1,\ldots,T,$$

$$(8) \quad (s^{n+1},z_t^{n+1}) = (s^*,\bar{z}_t), \quad (s^{n+2},z_t^{n+2}) = (s_*,\bar{z}_t) \qquad t=1,\ldots,T,$$

(9) $V_t^{n+1} = V_t^{n+2} = \alpha_t$ $\qquad t=1,\ldots,T,$

(10) $\theta \in \Theta$.

Theorem 3 suggests a procedure to find the set of maximum likelihood estimators for (V^*,θ^*): First find the set of solutions $(V_1^1,\ldots,V_T^1;\ldots,V_1^n,\ldots,V_T^n;\theta)$ to the constrained maximization problem described by (4)-(10) and second, find the set of functions V in W satisfying $V(s^i,z^i) = (V_1^i,\ldots,V_T^i)$ $(i=1,\ldots,n,n+1,n+2)$ for some solution vector $(V_1^1,\ldots,V_T^1;\ldots,V_1^n,\ldots,V_T^n;\theta)$.

If one wishes to obtain a single function in this set of solutions, it is only necessary to interpolate between the V_j^i values obtained. For example, one function that belongs to W and interpolates between the obtained values is the function $V(s,z)=(V_1(s,z),\ldots,V_T(s,z))$ defined by

$$V_t(s,z) = \min\{ V_t^i \mid (s,z_t) \leq (s^i,z_t^i),\ i = 1,\ldots,n,n+1,n+2\}.$$

We next note some computational properties of the maximization problem described in Theorem 3. First, the number of variables over which the maximization takes place is $n \cdot T$ plus the dimension of θ; in particular, the number of variables increases with the number of available observations. Second, the number of constraints is random; it depends on the partial ordering generated by $\leq$ on each of the T_t sets. The number of constraints corresponding to each of these sets is minimal when the domain of V_t^* is in the real line, in which case the number of these constraints can not exceed $n + 2$. Third, Assumption P.2 implies that the function in (4) is continuous in the variables $(V_1^1,\ldots,V_T^n)$ and θ. Fourth, when the probabilities satisfy (1) and $d_t^i=1$ the value of the function in (4) depends on the value of the differences $(V_t^i-V_1^i,\ldots,V_t^i-V_T^i)$ rather than on the value of $(V_1^i,\ldots,V_T^i)$ for $i=1,\ldots,n$. Fifth, as it is shown below, it is possible in certain situations to decrease considerably the number of variables and constraints over which the maximization takes place.

When the domain of one of the functions V_t is in the real line, it may be possible to group the variables V_t^i $(i=1,\ldots,n)$, reducing considerably the number of variables and constraints. Suppose, for example, that $T = J$, V_1 is defined on the real line, and for each j P_j is strictly increasing on the value of V_j and strictly decreasing on the value of V_k for $k\neq j$. In this situation, we will be able to group the variables V_1^i. All variables in the same group will attain the same value at any optimal solution of the constrained maximization problem, and hence can be considered as one variable. The monotonicity constraints need then only be imposed between consecutive groups.

To determine the groups of the set $\{ V_1^1,\ldots,V_1^n,V_1^{n+1},V_1^{n+2} \}$ of the above example, first order the vectors (s^i,z_1^i) $(i=1,\ldots,n+2)$ from smaller to larger. Next, include all variables to the right of any variable in the same group until the first V_1^i for which $d_1^i=0$, $d_1^r=1$, and $(s^r,z_1^r)\in F_1(s^i,z_1^i)$. The group ends at such V_1^i and a new group begins at V_1^r. To see why all the variables included in these groups will attain the same values at any optimal solution, note that the assumption that P_j is strictly increasing in V_j and strictly decreasing in V_k $(k\neq j)$ implies that if $d_1^i=1$ ($d_1^i=0$) the value of (4) will increase the bigger (smaller) the value of V_1^i. Hence, the optimal value of V_1^i will be the upper (lower) bound determined by the constraints. Since V_1 is defined on the real line, (s^i,z_1^i) possesses a unique immediate follower, say (s^r,z_1^r), and it is the immediate follower of exactly one point, say (s^t,z_1^t). Hence, if $d_1^i=1$ ($d_1^i=0$), the constraint $V_1^i \leq V_1^r$ ($V_1^t \leq V_1^i$) will be binding at any optimal solution. We can then include V_1^i and V_1^r (V_1^t and V_1^i) in the same group. The solution to the original optimization problem is identical then to the solution of the optimization problem in which all the variables in a same group are constrained to possess the same value; the monotonicity constraints need then be imposed only between one of the variables in each group.

Several methods exist for solving linearly constrained optimization problems of the kind described in Theorem 3. When the problem involves a large number of constraints, however, penalty methods seem to be the most appropriate, since they require less computer memory than other methods. Penalty methods transform a constrained maximization problem into a sequence of unconstrained maximization problems. The objective function in each of the unconstrained problems depends on the objective function of the constrained problem and on the constraint functions. The value of the objective function of the unconstrained problem is directly related to the value of the objective function of the constrained problem and indirectly related to the values of the constraint functions that do not satisfy the constraint. The penalty for being outside the constraint set increases the further away in the sequence of unconstrained problems the objective function is. Formally, to find by a penalty method a solution to the problem:

$$\begin{array}{ll} \text{Maximize}_x & f(x) \\ \text{subject to} & h(x) \le 0 \end{array}$$

where $f:R^L \to R$, $h:R^L \to R^N$, and $x \in R^L$, we solve the sequence, indexed by t, of unconstrained maximization problems:

$$\text{(15) Maximize}_x \quad f(x) - K_t\, P(x)$$

where $K_t > 0$, $K_t \to \infty$, and P(x) is a continuous function that satisfies $P(x) \ge 0$, for all $x \in R^L$, and $P(x) = 0$ if and only if $h(x) \le 0$.

For small enough K_0, the maximum of (15) will be relatively close to the unconstrained maximum of $f(\cdot)$. As $K_t \to \infty$, the point at which (15) is maximized approaches a point in the constraint set. The point at which the t-th problem is maximized is employed as the starting point for the maximization of the t+1-th problem. If $f(\cdot)$ is continuous, any limit point of the sequence of solutions of the unconstrained problems will be a solution to the constrained problem (see Fiacco and McCormick (1968) for more details about the theory and practice of this procedure.)

9. CONCLUSION

We have presented an estimation method that requires weak monotonicity assumptions about a nonparametric function, V. The vector-valued function V was only required to be monotone increasing, upper-semicontinuous, and uniformly bounded. The estimation method was developed for multinomial models. The probability of each outcome was assumed to be known up to a finite dimensional parameter vector and to be a function of V. Polychotomous choice models are among the models to which the new estimation method can be applied.

We have given conditions under which both the function V and the parameter of the probability functions are identified. A maximum likelihood method to obtain an estimator for V and the parameter of the probability functions was described and the consistency of the estimator was established.

We have described a method of implementing the computation of these estimators. This method proceeded by transforming the problem of maximizing the likelihood function over a function set into the problem of maximizing the likelihood function over a finite number of variables subject to linear inequalities. We have shown that the number of variables and constraints can be decreased considerably in certain situations. The use of penalty function methods to solve constrained optimization problems of the kind described in this paper was discussed.

Comparison of our theoretical results with those obtained in Matzkin (1989), for polychotomous choice models, implies that it is possible to impose weaker assumptions about the subutility function than those imposed in Matzkin (1989) and still be able to identify the model. The consistency of the new estimator, however, is only obtained with respect to a metric that is weaker than the one employed in Matzkin (1989).

Our results also show that it is possible to "invert" the parametric restrictions in the distribution of the random term and the systematic subutility, which were made in the binary choice model of Cosslett (1983). That is, instead of assuming that the subutility function is parametric and the distribution of the random terms is monotone increasing, it is possible to assume that the distribution of the random terms is parametric and the subutility function is monotone increasing. This new specification allows us to estimate semiparametric choice models in which the number of alternatives is larger than two and the distribution of the random terms depends on the exogenous variables. Moreover, the new estimators are obtained by maximizing a continuous, typically well behaved, function, instead of a discontinuous function.

APPENDIX

PROOF OF THEOREM 1: Let $(V,\theta)\in(W\times\Theta)$ be such that $(V,\theta)\neq(V^*,\theta^*)$. We will show that then, for some $j \in \{1,\ldots,J\}$ and some $(s^*,z^*)\in S\times Z$

(1.1) $P(j|s^*,z^*;V,\theta) \neq P(j|s^*,z^*;V^*,\theta^*)$.

If $\theta\neq\theta^*$, this follows by Assumption P.5.

If $\theta=\theta^*$ then it must be that for some $(s,z)\in S\times Z$ and $t\in\{1,\ldots,T\}$ $V_t^*(s,z) \neq V_t(s,z)$. Let z^* be such that $z_t^*=z_t$ and $z_k^*=\bar{z}_k$ $(k\neq t)$, where $\bar{z}$ is as given by Assumption W.4. Then, it follows that for all $k\neq t$

$$V_k^*(s,z^*) = V_k^*(s,\bar{z}) \quad \text{and} \quad V_k(s,z^*) = V_k(s,\bar{z})$$

by Assumption W.3; and

$$V_k^*(s,\bar{z}) = V_k(s,\bar{z}) = \alpha_k$$

by Assumption W.4. From Assumption W.3 it also follows that

$$V_t^*(s,z^*) = V_t^*(s,z) \quad \text{and} \quad V_t(s,z^*) = V_t(s,z).$$

Hence, $V_t^*(s,z^*) \neq V_t(s,z^*)$ and for all $k\neq j$ $V_k^*(s,z^*) = V_k(s,z^*)$.

By Assumption P.3 it then follows that for some $j\in A$

$$P(j|s^*,z^*;V,\theta) \neq P(j|s^*,z^*;V^*,\theta^*).$$

Hence, (1.1) follows.

From Assumptions P.1 and P.2 it now follows that there exists some $\mu > 0$ such that

(1.2) $\|(s,z) - (s^*,z^*)\| < \mu$, $\|V(s,z) - V(s^*,z^*)\| < \mu$, and
$\|V^*(s,z) - V^*(s^*,z^*)\| < \mu$

imply that

(1.3) $P(j|s,z;V,\theta) \neq P(j|s,z;V^*,\theta^*)$.

The proof of Theorem 1 will then be completed if we show that for some subset $Y(V,\theta)$ of the μ-neighborhood of (s^*,z^*), $Y(V,\theta)$ possesses positive probability and for all $(s,z)\in Y(V,\theta)$, (1.3) is satisfied.

To show the existence of such set $Y(V,\theta)$, we will prove that the upper semicontinuity of V implies that there exist neighborhoods $N^*(V)$ and $N^*(V^*)$ of (s^*,z^*) such that for all $(s,z)\in(N^*(V)\cap N(V))$ and all $t\in\{1,\ldots,T\}$

(1.4) $\quad V_t(s,z) \le V_t(s^*,z^*) + \mu \quad$ and $\quad V_t^*(s,z) \le V_t^*(s^*,z^*) + \mu$.

Suppose that $\forall\ \delta_n > 0$ there exists (s_n,z_n) such that $\|(s_n,z_n)-(s^*,z^*)\| < \delta_n$ and $V_t(s_n,z_n) > V_t(s^*,z^*) + \mu$. Let $\delta_n \to 0$. Then, $(s_n,z_n) \to (s^*,z^*)$ and for all n, $V(s_n,z_n) \le V(s^*,z^*) + \mu$. Since V is upper-semicontinuous, this implies that $V(s^*,z^*) \ge V(s^*,z^*) + \mu$, which is a contradiction. Hence, there exists $N^*(V)$ such that $\forall(s,z)\in N^*(V)$ $\quad V_t(s,z) \le V_t(s^*,z^*) + \mu$

Employing a similar argument for V^*, we can conclude that there exists a neighborhood $N^*(V^*)$ of (s^*,z^*) such that for all $(s,z)\in N^*(V^*)$ $\quad V^*(s,z) \le V^*(s^*,z^*) + \mu$. Hence, (1.4) is proved.

Let

$$Y(V,\theta) = N^*(V) \cap N^*(V^*) \cap \{(s,z) \mid (s,z) \le (s^*,z^*)\} \cap N((s^*,z^*),\eta),$$

where $N((s^*,z^*),\eta)$ is a η-neighborhood of (s^*,z^*). Then, the monotonicity of V and V^* (Assumption W.1) and the definition of $Y(V,\theta)$ imply that for all $(s,z) \in Y(V,\theta)$ and all t

$0 \le V_t(s,z) - V_t(s^*,z^*) \le \mu$, $0 \le V_t^*(s,z) - V_*^t(s^*,z^*) \le \mu$, and $\|(s,z)$ $(s^*,z^*)\| < \mu$. Hence, for all $(s,z) \in Y(V,\theta)$

$$P(j|s,z;V,\theta) \ne P(j|s,z;V^*,\theta^*)\ .$$

This completes the proof of Theorem 1.

To prove Theorem 2, we follow Matzkin (1989) and let $f(x;V,\theta)$ denote the probability density of x when $(V^*,\theta^*) = (V,\theta)$. Hence,

$$f(x;V,\theta) = g(s,z) \prod_{j=1}^{J} [P(j|s,z;V,\theta)]^{d_j} .$$

The probability measure of $f(x;V^*,\theta^*)$ will be denoted by P^*, the set $\{ (d_1,\ldots,d_t) \mid d_t \in \{0,1\},\ \Sigma_{j=1}^{J} d_t = 1 \}$ will be denoted by D, and the set $D \times S \times Z$ will be denoted by X. We next prove some lemmas.

LEMMA 1: *$[(W\times\Theta),d]$ is a compact metric space.*

PROOF: We first show that (W,m) is a metric space. It is clear that for all $V,V',\bar{V} \in W$, $m(V,V')=m(V',V)$, $m(V,V') \le m(V,\bar{V}) + m(\bar{V},V)$, and $V=V'$ implies $m(V,V') = 0$. Suppose that $m(V,V') = 0$ and $V \neq V'$. Then, there exists a subset $C \subset S\times Z$ possessing positive Lebesgue measure and such that for all $r\in C$, $V(r) \neq V'(r)$. Since the set of points that are points of discontinuity of either V' or V possesses Lebesgue measure zero (see, e.g. Matzkin and Meyers (1986)), there exists $r' \in C$ at which both V and V' are continuous. Hence, since $V(r') \neq V'(r')$, there exists a neighborhood N of r' such that $\forall\ r\in N$, $V(r) \neq V'(r)$. But this implies that $m(V,V') > 0$, which is a contradiction. Hence, $m(V,V')=0$ implies that $V=V'$. We have then shown that (W,m) is metric space.

We next show that (W,m) is compact. Since it is a metric space, it suffices to show that (W,m) is sequentially compact. Let then $\{V_i\}$ be a sequence in W. By the standard diagonalization principle (see, e.g. the proof of Helly's Theorem), there exists a countable dense subset Q of $S\times Z$, a subsequence $\{V_k\}$ of $\{V_i\}$, and a function $\bar{V}:Q \to R^T$ such that for all $q \in Q$, $V_k(q) \rightarrow V(q)$. Let $V': S \times Z \rightarrow R^T$ be defined by $V'(r) = \inf \{ V(q) \mid r \leq q , q \in Q \}$. Then, V' is monotone increasing, for all $r \in S \times Z$ $L \leq V'(r) \leq U$, and $\{V_k\}$ converges to V' at all points of continuity of V'. Define $V: S \times Z \rightarrow R^T$ by $V(r) = \lim_{n\rightarrow\infty} \sup\{ V'(r') \mid \|r'-r\| < (1/n) \}$. Then, V is upper-semicontinuous, monotone increasing, and for all $r \in S \times Z$ $L \leq V(r) \leq U$. Hence, $V \in W$.

It only remains to show that $m(V_k, V) \rightarrow 0$.

Since for all $r \in S \times Z$ at which V' is continuous, $V'(r) = V(r)$, and since the sequence V_k converges pointwise to V' at all points of continuity of V', it follows that V_k converges pointwise to V at all points of continuity of V'. Then, since the set of points of continuity of V' has Lebesgue measure zero, V_k converges to V a.e. That $m(V_k, V) \rightarrow 0$ then follows by the uniform boundedness of V and the V_{i_k} functions, the definition of m, and Lebesgue Dominated Convergence Theorem. This concludes the proof that W is sequentially compact with respect to m.

That $((W \times \Theta), d)$ is a compact metric space now follows from the above result and the assumption that $(\Theta, \|\cdot\|)$ is a compact metric space.

LEMMA 2 (continuity of probability densities on $(W \times \Theta)$:) *Except perhaps for a subset of X possessing zero probability, $f(x;V,\theta)$ is continuous on $(W \times \Theta)$.*

PROOF: It suffices to show that for all $j \in A$ $P(j|s,z;V,\theta)$ is continuous in (V,θ) a.s. (G). Let then $\{(V^n,\theta^n)\}_{n=1}^{\infty} \subset (W \times \Theta)$ and $(V,\theta) \in (W \times \Theta)$ be such that $\lim_{n\rightarrow\infty} d[(V^n,\theta^n),(V,\theta)]=0$. Then, $m(V^n,V) \rightarrow 0$ and $\|\theta^n - \theta\| \rightarrow 0$. Suppose that $m(V^n,V) \rightarrow 0$ implies that V^n converges pointwise to V at all points of continuity of V. Then, by Assumption P.2, $P(j|s,z;V,\theta)$ is continuous at (V,θ) if (s,z) is a point a continuity of V. Since the set of points of discontinuity of V has zero Lebesgue measure, and G is absolutely continuous, it follows that except perhaps for a set that possesses zero probability, $P(j|s,z;V,\theta)$ is continuous on $(W\times\Theta)$. Hence, the lemma will be shown if we prove that $m(V^n,V) \to 0$ implies that for all $(s,z)\in S\times Z$ at which V is continuous and all t, $V_t^n(s,z) \to V_t(s,z)$. We proceed to show this claim. Suppose that V is continuous at (s,z) but for some t, some $\mu > 0$, and some subsequence $\{V_t^k\}$ of $\{V_t^n\}$, $|V_t^k(s,z)-V_t(s,z)| > \mu$. Divide the sequence $\{V_t^k\}$ into the subsequences $\{V_t^{k+}\}$ and $\{V_t^{k-}\}$ such that for all k+ and all k-

$$V_t^{k+}(s,z) > V_t(s,z) + \mu \quad \text{and} \quad V_t^{k-}(s,z) < V_t(s,z) + \mu.$$

Since V is continuous at (s,z) and the V^k functions are monotone increasing, there exists $\delta > 0$ such that for all (s',z') in the δ-neighborhood, $N((s,z),\delta)$, of (s,z), $|V(s',z') - V(s,z)| < (\mu/2)$. Let $A = \{(s',z')\in N((s,z),\delta)|\ (s',z')\geq(s,z)\}$ and let $B = \{(s',z')\in N((s,z),\delta)|\ (s',z')\leq(s,z)\}$. Then, by the monotonicity of the V^k functions,

$$V_t^{k+}(s',z') \geq V_t^{k+}(s,z) > V_t(s,z) + \mu \geq V_t(s',z') + (\mu/2)$$

if $(s',z')\in A$ and

$$V_t^{k-}(s',z') \leq V_t^{k-}(s,z) < V_t(s,z) - \mu \leq V_t(s',z') - (\mu/2)$$

if $(s',z')\in B$. Since the sets A and B possesses positive Lebesgue measure, the above implies that $m(V^{k+},V)$ and $m(V^{k-},V)$ are both uniformly bounded away from zero, which is impossible since $m(V^k,V) \to 0$. Hence, if V is continuous at (s,z), $V^n(s,z) \to V(s,z)$. This completes the proof of the lemma.

LEMMA 3 (measurability:) *Define the function* $f':(X\times W\times\Theta\times R_{++}) \to R$ *by* $f'(x,V,\theta,\varepsilon) = \sup_{(V',\theta')\in(W\times\Theta)} \{f(x,V',\theta') | d[(V,\theta),(V',\theta')] < \varepsilon\}$. *Then, for all* (V, θ) *and for small enough* $\varepsilon > 0$, f' *is measurable in* x.

PROOF: Since by Lemma 1 $((W\times\Theta),d)$ is a compact metric space, there exists a countable dense subset $(W'\times\Theta')$ of $(W\times\Theta)$. For any $x\in X$ define

$t = \sup\{f(x;V',\theta') \mid d[(V',\theta'),(V,\theta)] < \varepsilon,\ (V',\theta') \in (W \times\Theta)\}$ and

$r = \sup\{f(x;V_i,\theta_i) \mid d[(V_i,\theta_i),(V,\theta)] < \varepsilon,\ (V_i,\theta_i) \in (W'\times\Theta')\}$.

We will show that $r = t$. Since $(W'\times\Theta') \subset (W\times\Theta)$, $r \le t$. Suppose that $r < t$. Then, there exists $(V,\theta)\in(W\times\Theta)$ such that

(3.1) $\quad f(x;V,\theta)) > f(x;V_i,\theta_i) \quad$ for all $(V_i,\theta_i)\in (W'\times\Theta')$.

Let $\eta > 0$ be sufficiently small and for each $t\in\{1,\ldots,T\}$, let w_t denote the vector $(s,z_t) - (\eta,\ldots,\eta)$. Consider the function V' defined by

$V_t'(s',z_t') = V_t(s',z_t')$ if $V_t(s',z_t') \ge V_t(s,z_t)$,

$V_t'(s',z_t') = V_t(s,z_t)$ if $V_t(s',z_t') < V_t(s,z_t)$ and $w_t \le (s',z_t')$, and

$V_t'(s',z_t') = V_t(s',z_t')$ otherwise.

Then, $V'\in W$, V' is continuous at (s,z), and since $(V_1(s,z),\ldots,V_T(s,z) = (V_1'(s,z),\ldots,V_T'(s,z))$, $f(x;V',\theta') = f(x;V,\theta)$. Since W' is dense in W, there exists a sequence $\{V_i\} \subset W'$ such that $m(V_i,V') \to 0$. But, since V' is continuous at (s,z), this implies, by the argument given in the proof of Lemma 2, that $V_i(s,z) \to V(s,z)$. The continuity of $\bar{P}(j|s,z;V_1,\ldots,V_T,\theta)$ in $(V_1,\ldots,V_T,\theta)$ (Assumption P.2) implies then that $f(x;V_i,\theta_i) \to f(x;V',\theta) = f(x;V,\theta)$, which contradicts (3.1). It then follows that $f(x;V,\theta)$ is measurable in x if for each $i=1,2,\ldots$ $f(x;V_i,\theta_i)$ is measurable in x. But this follows easily because $g(\cdot)$ is measurable, $(V_1(s,z),\ldots,V_T(s,z))$ is measurable in (s,z), and for each $j\in A$ $P(j|s,z;V,\theta)$ is continuous in (s,z) and the vector $(V_1(s,z),\ldots,V_T(s,z))$.

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

PROOF: See the proof of Lemma 6 in Matzkin (1989).

LEMMA 5: *Define the function* $f^*:(X \times W \times \Theta \times R_{++}) \to R$ *by*

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

Then, for any $(V,\theta) \in (W \times \Theta)$ *and for sufficiently small* $e > 0$ $\int_X \log f*(x,V,\theta,e) \, dP^*(x)$ *is finite.*

PROOF: See the proof of Lemma 7 in Matzkin (1989).

LEMMA 6 (identification:) *If* $(V,\theta) \in (W \times \Theta)$ *and* $(V,\theta) \neq (V^*,\theta^*)$ *then for some set* $E \subseteq X$ *with* $P^*(E) > 0$, $\int_E f(x;V,\theta) \, dx \neq \int_E f(x;V^*,\theta^*) \, dx$.

PROOF: By Theorem 1, there exists $j \in A$ and $Y(V,\theta) \subset S \times Z$ such that $G(Y(V,\theta)) >$ and for all $(s,z) \in Y(V,\theta)$

$$P(j|s,z;V,\theta) \neq P(j|s,z;V^*,\theta^*).$$

Let $E = \{ (d,s,z) \in D \times S \times Z \mid d_t = 1, (s,z) \in N \}$. Since by Assumption P.4 $P(j|s,z;V^*,\theta^*) > 0$ for all $(s,z) \in Y(V,\theta)$, $P^*(E) > 0$. Hence,

$$\int_E f(x; V, \theta) \, dx = \int_{N \cap (S \times Z))} P(j|s,z,V,\theta) \quad dG(s,z)$$
$$\neq \int_E f(x;V^*,\theta^*) \, dx = \int_{N \cap (S \times Z))} P(j|s,z,V^*,\theta^*) \, dG(s,z) \; .$$

PROOF OF THEOREM 2: The result follows from Lemmas 1-6 by the argument given in Matzkin (1989).

PROOF OF THEOREM 3: We first show that the set of vectors $(V_1^1,\dots,V_T^1;\dots;V_1^n,\dots,V_T^n)$ for which there exists some $V\in W$ with $V(s^i,z^i)=(V_1^i,\dots,V_T^i)$ $(i=1,\dots,n)$ is the set of vectors $(V_1^1,\dots,V_T^1;\dots;V_1^n,\dots,V_T^1)$ that satisfy (5)-(9): It is clear that if $(V,\theta)\in(W\times\Theta)$, the vector $(V_1(s^1,z^1),\dots,V_T(s^1,z^1);\ \dots\ ;V_1(s^n,z^n),\dots,V_T(s^n,z^n);\ \theta)$ satisfies (5)-(10). Suppose then that $(V_1^1,\dots,V_T^1;\dots;V_1^n,\dots,V_T^n)$ satisfies (5)-(9). For each t, define the function V_t on the set of all $(s,z_t)\le(s^*,z_t^*)$ by

$$V_t(s,z_t) = \min\{\, V_t^i \mid (s,z_t)\le(s^i,z_t^i)\quad i=1,\dots,n+1,n+2 \,\}\ (t=1,\dots,T).$$

Then, by (5) V_t is monotone, by definition V_t is upper-semicontinuous, and by (6) and (7) V_t is uniformly bounded by L_t and U_t. To show that for all $t\in\{1,\dots,T\}$ and all $s\le s^*$, $V_t(s,\bar{z}_t)=\alpha_t$, we note that the definitions of s^*, s_*, $\bar{z}_t$, and V_t imply that for all s such that $s\le s_*$ $V_t(s,\bar{z}_t)=V_t(s_*,\bar{z}_t)=\alpha_t$ and for all s such that $s_*\le s\le s^*$, $\alpha_t=V_t(s_*,\bar{z}_t)\le V_t(s,\bar{z}_t)\le V_t(s^*,\bar{z}_t)=\alpha_t$.

Extend now the function V_t to all of $S\times Z_t$ by letting

$V_t(s,z_t) = \alpha_t$ if (s,z_t) is not bigger than (s^*,z_t^*) and $z_t\le\bar{z}_t$

and

$V_t(s,z_t) = U_t$ if (s,z_t) is not bigger than (s^*,z_t^*) and z_t is not smaller than $\bar{z}_t$.

Then, the function V whose coordinates are the functions V_t so defined belongs to the set W and its value at (s^i,z^i) is $(V_1^i,\dots V_T^i)$ $(i=1,\dots,n)$.

We have then shown that the set of vectors whose coordinates are the values of a function in W at the observed points can be characterized by the set of vectors satisfying (5)-(9). Since, by Assumption P.1, (2) depends only on the values of V at the observed points, maximization of (2) over W is then equivalent to maximization of (4) over the set of vectors satisfying (5)-(9).

Q.E.D.

REFERENCES

ABRAMAZAR, A. and P. SCHMIDT (1982), "An Investigation of the Robustness of the Probit Estimator to Non-Normality," *Econometrica*, Vol. 50, No. 4.

BARLOW, R.E., D.J. BARTHOLOMEW, J.M. BREMNER, and H.D. BRUNK (1972), *Statistical Inference Under Order Restrictions*, London: John Wiley and Sons.

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

FIACCO, A.V and G.P. McCORMICK (1968), *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, Inc.

HAN, A.K. (1987), "Nonparametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimation," *Journal of Econometrics*, 35, 303- 316.

HAUSMAN J.A. and D.A. WISE (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, Vol. 46, 403-26.

ICHIMURA, H. (1988), "Estimation of Single Index Models," University of Minnesota.

KLEIN, R. W. and R.H. SPADY (1988), "Semiparametric Estimation of Discrete Choice Models," Bell Communication Research.

MANSKI, C.F. (1975), "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205-228.

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

________. (1988), "Identification of Binary Response Models", *Journal of the American Statistical Association*, Vol.83, No. 403.

MATZKIN, R.L. (1988), "Nonparametric and Distribution-Free Estimation of Threshold Crossing and Binary Choice Models," Cowles Foundation Discussion Paper No. 889, Yale University.

________. (1989), "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," Cowles Foundation, Yale University.

________. (1990a), "Least-concavity and the Distribution-free Estimation of Observationally-equivalent Concave Functions," Cowles Foundation, Yale University.

________. (1990b), "A Nonparametric Maximum Rank Correlation Estimator", in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, by W. Barnett, J. Powell, and G. Tauchen (eds.), Cambridge: Cambridge University Press, forthcoming.

MATZKIN, R.L. and N. MEYERS (1986), "Monotone Functions," mimeo, University of Minnesota.

McFADDEN, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Mathematical Methods in the Social Sciences*, by P. Zarembka (ed.), New York: Academic Press.

________. (1981), "Econometric Models of Probabilistic Choice", Chapter 5 in *Structural Analysis of Discrete Data with Econometric Applications*, by C.F. Manski and D. McFadden (eds.), The MIT Press.

POWELL, J.L., J.H. STOCK, and T.M. STOKER (1990), "Semiparametric Estimation of Index Coefficients," *Econometrica*.

RAO, PRAKASA (1983), *Nonparametric Functional Estimation*, Academic Press.

ROBERTSON, T., F.T. WRIGHT, and R.L. DYKSTRA (1988), *Order Restricted Statistical Inference*, John Wiley and Sons.

STOKER, T.M. (1986), "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461-1481.

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

NOTES

* The support of National Science Foundation through Grant SES-8720596 is gratefully acknowledged. I am thankful to Alvin Klevorick, Lung-Fei Lee, and Christopher Sims for their comments and suggestions.

1. A function $V:R^K \to R^T$ is *monotone increasing* if for all $x,y \in R^K$ such that $x_k \geq y_k$ $(k=1,..,K)$, $V_t(x) \geq V_t(y)$ $(t=1,...,T)$; $V:R^K \to R^T$ is *upper-semicontinuous* if for all $\alpha \in R^T$ the set $\{x \in R^K \mid V_t(x) \geq \alpha_t$ for $t=1,...,T\}$ is a closed set; $V:R^K \to R^T$ is *continuous* if for all $\alpha \in R^T$ the sets $\{x \in R^K \mid V_t(x) \geq \alpha_t$ for $t=1,..,T\}$ and $\{x \in R^K \mid V_t(x) \leq \alpha_t$ for $t=1,...,T\}$ are closed; $V:R^K \to R^T$ is *uniformly bounded* if there exist $\alpha, \beta \in R^T$ such that for all $x \in R^K$ and all $t=1,...,T$ $\alpha_t \leq V_t(x) \leq \beta_t$.

2. The estimation of monotone density, distribution, median, and mean functions has been extensively studied in the statistics literature. (For surveys of this literature, see, Barlow, Bartholomew, Bremner, and Brunk (1972), Prakasa Rao (1983), and Robertson, Wright, and Dykstra (1988).)

3. We say that $V=V'$ whenever V and V' are equal at every point except, perhaps, for a set of Lebesgue measure zero.

4. In the particular case in which $T = J$ and the probabilities of the outcomes are generated according to (1), it is possible to weaken

Assumptions W.3 and W.4 at the cost of strengthening Assumption P.3 (see Matzkin (1989)).

5. The metric m is weaker than the essential supremum metric employed in Matzkin (1989).

6. These probabilities can be generated by a polychotomous choice model in which the random vector $\varepsilon=(\varepsilon_1^i,\varepsilon_2^i,\varepsilon_3^i)$ possesses a Generalized Extreme Value distribution of the form

$$\Pr(\varepsilon \le \eta) = \exp\left[-\left[y_1 + (y_2^{1/(1-\theta)} + y_3^{1/(1-\theta)})^{(1-\theta)} \right] \right]$$

where $\eta=(\eta_1,\eta_2,\eta_3)$, $y_1 = \exp(-\eta_1)$, $y_2 = \exp(-\eta_2)$, and $y_3 = \exp(-\eta_3)$.

7. To see this, let $w_i = \exp(V_i)$ $(i=1,2,3)$ and $c = (w_2^{1/(1-\theta)} + w_3^{1/(1-\theta)})^{(1-\theta)}$. When $z = 0$, $w_i = 1$ $(i=1,2,3)$ and $c = 2$. For any two differentiable functions $f(x)$ and $g(x)$, the derivative of $f(x)^{g(x)}$ with respect to x is

$$f(x)^{g(x)} \left[g'(x) \ln(f(x)) + g(x) (f'(x)/f(x)) \right],$$

where $f'(x)$ and $g'(x)$ denote, respectively, the derivative of f and g with respect to x. Hence, the derivative of $P(1|z;\, V,\theta)$ with respect to θ is given by

$$-\frac{w_1 \quad c^{(1-\theta)}}{[\, w_1 + c^{(1-\theta)}\,]^2} \left\{ -\ln(c) + \frac{1}{c} \left[\frac{w_2^{1/(1-\theta)} \ln(w_2) + w_3^{1/(1-\theta)} \ln(w_3)}{(1-\theta)} \right] \right\}$$

and it equals

$$\frac{\ln(2) \quad 2^{(1-\theta)}}{[\, 1 + 2^{(1-\theta)}\,]^2} < 0 \qquad \text{for all } \theta\in\Theta, \text{ when } z = 0.$$