

The copyright to this article is held by the Econometric Society, <http://www.econometricsociety.org/>. It may be downloaded, printed and reproduced only for personal or classroom use. Absolutely no downloading or copying may be done for, or on behalf of, any for-profit commercial firm or for other commercial purpose without the explicit permission of the Econometric Society.

All other permission requests or questions (including commercial purposes or on behalf of any for-profit entity) should be addressed to:

Wiley Permissions

www.wiley.com/go/rightslicensing

Then select Copyright & Permissions

For requests for any content not containing a 'request permission' link please contact permissions@wiley.com .

SEMPARAMETRIC ESTIMATION OF MONOTONE AND CONCAVE UTILITY FUNCTIONS FOR POLYCHOTOMOUS CHOICE MODELS

BY ROSA L. MATZKIN¹

This paper introduces a semiparametric estimation method for polychotomous choice models. The method does not require a parametric structure for the systematic subutility of observable exogenous variables. The distribution of the random terms is assumed to be known up to a finite-dimensional parameter vector. In contrast, previous semiparametric methods of estimating discrete choice models have concentrated on relaxing parametric assumptions on the distribution of the random terms while leaving the systematic subutility parametrically specified.

The systematic subutility is assumed to possess properties, such as monotonicity and concavity, that are typically assumed in microeconomic theory. The estimator for the systematic subutility and the parameter vector of the distribution is shown to be strongly consistent. A computational technique to calculate the estimators is developed.

KEYWORDS: Semiparametric, discrete choice, concavity, monotonicity, maximum likelihood, strong consistency.

1. INTRODUCTION

THIS PAPER INTRODUCES a semiparametric method of estimating polychotomous choice models that does not require a parametric specification for the systematic subutility function. Instead of positing a parametric structure, I assume that the subutility possesses properties, such as concavity and monotonicity, that are consistent with assumptions typically made in microeconomic theory. The distribution of the random terms, which may depend on the exogenous variables, is assumed to be known up to a finite-dimensional parameter vector.² As long as the assumptions are satisfied, the local behavior of the systematic subutility is likely to be better uncovered by this method than by methods that impose a parametric structure on this function.

The estimator is obtained by maximizing the likelihood function over a set of nonparametric functions W and a finite dimensional parameter set Θ . The functions in W possess the same properties that the systematic function is assumed to possess. The set Θ is assumed to contain the value of the parameter of the distribution of the random terms. To compute the maximum likelihood estimator, the maximization over the function space is transformed into a large

¹ I am indebted to Lung-Fei Lee, Marcel K. Richter, and Christopher Sims for their advice. I also thank Philip Dybvig, Vassilis Hajivassiliou, Alvin Klevorick, three anonymous referees, and a co-editor, for their useful comments and suggestions. The support of NSF through Grants SES-8720596 and SES-8900291 is gratefully acknowledged.

² The estimation of discrete choice models in which the distribution of the random terms is nonparametric but the systematic subutility is parametric has been studied by Manski (1975, 1985), Cosslett (1983), Stoker (1986), Han (1987), Ichimura (1988), and Klein and Spady (1988), among others. The estimation of discrete choice models in which neither the systematic function nor the distribution of the random term is parametric has been recently studied by Matzkin (1990, 1992).

linearly-constrained maximization problem. The estimator is shown to be strongly consistent by adapting the consistency result of Wald (1949).

The model and the estimator are presented in Section 2. Section 3 shows the strong consistency result, and Section 4 describes the technique to compute the estimator. Section 5 summarizes the conclusions. The proofs of the theorems and lemmas are presented in the Appendix.

2. THE POLYCHOTOMOUS CHOICE MODEL

In this model, a typical consumer chooses a single alternative from a finite set A of J alternatives. The consumer is assumed to choose the alternative that maximizes his utility. The utility of each alternative j in A is the sum of a subutility function $V_j^*(\cdot)$ and an unobservable random term ε_j . For each alternative j , $V_j^*(\cdot)$ is a function of observable socioeconomic characteristics s of the consumer and observable attributes z_1, \dots, z_J of the $1, \dots, J$ th alternative.

The random vector $(s, z) = (s, z_1, \dots, z_J)$ will be assumed to possess a probability density $g(\cdot)$ whose support will be denoted by $S \times Z$, where $Z = \prod_{j=1}^J Z_j$. The probability measure of (s, z) will be denoted by $G(\cdot)$. The vector $\varepsilon = (\varepsilon_1, \dots, \varepsilon_J)$ will be assumed to possess, conditional on (s, z) , a Lebesgue density $q(\varepsilon; s, z, \theta^*)$ that is known up to the finite dimensional parameter vector θ^* . The probability that, given observable characteristics (s, z) , a consumer will choose an alternative $t \in A$, will be denoted by $P(t|s, z; V^*, \theta^*)$. Our assumptions imply that

$$\begin{aligned}
 (1) \quad P(t|s, z; V^*, \theta^*) &= \text{Prob} \{V_t^*(s, z) + \varepsilon_t \\
 &\geq V_k^*(s, z) + \varepsilon_k; k \neq t, k = 1, \dots, J\} \\
 &= \int_{\varepsilon_t = -\infty}^{\infty} \int_{\varepsilon_1 = -\infty}^{V_t^* - V_1^* + \varepsilon_t} \dots \\
 &\int_{\varepsilon_J = -\infty}^{V_t^* - V_J^* + \varepsilon_t} q(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_J; s, z, \theta^*) d\varepsilon_1 \cdots d\varepsilon_J,
 \end{aligned}$$

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

This paper is concerned with the problem of estimating the subutility function V^* and the parameter vector θ^* from n independent observations. Each observation x consists of a vector (s, z) and a chosen alternative denoted by a vector $d = (d_1, \dots, d_J)$, where for each j d_j equals one if the alternative j was chosen and equals zero otherwise. The conditional log-likelihood function for the n independent observations $\{x^i\}_{i=1}^n$ is then

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

We define our *maximum likelihood estimator* of (V^*, θ^*) to be the pair $(V_n^{ML}, \theta_n^{ML})$ that maximizes (2) over a set $(W \times \Theta)$, where Θ is a subset of a

Euclidean space and W is a set of nonparametric functions $V: S \times Z \rightarrow R^J$ such that $(V^*, \theta^*) \in (W \times \Theta)$.

The Fourier flexible form of Gallant (1981, 1982) and Gallant and Golub (1984) could also be employed to estimate, by a maximum likelihood method, a nonparametric systematic function. These estimators can be restricted to be monotone, concave, or linearly homogeneous. In particular, the computation of the unconstrained estimators is, at the present moment, simpler than ours. Their consistency has been shown under stronger conditions than those required by our maximum likelihood estimator.³

3. CONSISTENCY OF THE MAXIMUM-LIKELIHOOD ESTIMATOR

The following assumptions will be made:

Let $m: W \times W \rightarrow R_+$ be a metric on W and define the metric d on $(W \times \Theta)$ by $d[(V', \theta'), (V'', \theta'')] = \|\theta' - \theta''\| + m(V', V'')$, where $\|\cdot\|$ denotes the Euclidean metric.

ASSUMPTION 1.1: Θ is a compact subset, with respect to $\|\cdot\|$, of R^L ($L < \infty$).

ASSUMPTION 1.2: $\theta^* \in \Theta$.

ASSUMPTION 2.1: W is a set of functions $V: S \times Z \rightarrow R^J$ that is compact with respect to m .

ASSUMPTION 2.2: $V^* \in W$.

ASSUMPTION 2.3: If $\{V_n\} \subset W$, $V \in W$, and $m(V_n, V) \rightarrow 0$, then for all $(s, z) \in S \times Z$ $\|V_n(s, z) - V(s, z)\| \rightarrow 0$.

ASSUMPTION 2.4: $\forall V \in W$, V is continuous on $S \times Z$.

ASSUMPTION 2.5: $\forall V \in W$, $\forall (s, z) \in S \times Z$, and $\forall k \neq 1$, $V_1(s, z)$ does not depend on z_k and $V_k(s, z)$ does not depend on z_1 .

ASSUMPTION 3.1: The support of G is $S \times Z$.

ASSUMPTION 3.2: The probability density g is uniformly bounded.

ASSUMPTION 4.1: Conditional on $(s, z) \in S \times Z$, for each $\theta \in \Theta$, ε possesses a conditional Lebesgue density $q(\varepsilon; s, z, \theta)$.

³Also relevant in this context is the statistics literature about estimation subject to shape restrictions. For a review of this literature, see Prakasa Rao (1983) and Robertson, Wright, and Dykstra (1988).

ASSUMPTION 4.2: *There exists a Lebesgue integrable function $\phi(\varepsilon)$ such that, for all $(s, z, \theta) \in S \times Z \times \Theta$, $|q(\varepsilon; s, z, \theta)| \leq \phi(\varepsilon)$ a.e. in ε .*

ASSUMPTION 4.3: *For a.e. ε , $q(\varepsilon; s, z, \theta)$ is continuous in (s, z, θ) .*

ASSUMPTION 4.4: *Conditional on $(s, z) \in S \times Z$, for all $\theta \in \Theta$, the support of the conditional density $q(\varepsilon; s, z, \theta^*)$ is R^J .*

ASSUMPTION 5: *There exist $(\alpha_1, \dots, \alpha_J) \in R^J$ and $(\bar{s}, \bar{z}) \in S \times Z$ such that the following assumptions hold.*

ASSUMPTION 5.1: $\forall V \in W$ and $\forall s \in S$, $V_1(s, \bar{z}) = \alpha_1$.

ASSUMPTION 5.2: $\forall V \in W$ and $\forall j \in A$, $V_j(\bar{s}, \bar{z}) = \alpha_j$.

ASSUMPTION 5.3: $\forall \theta \in \Theta$ such that $\theta \neq \theta^*$ there exists $j \in A$ such that $P(j|\bar{s}, \bar{z}; V^*, \theta^*) \neq P(j|\bar{s}, \bar{z}; V^*, \theta)$.

Assumptions 1.1 and 2.1 are employed as a substitute for the assumption made in Wald (1949) that the probability density of the observations converges to zero as the norm of the parameters tends to infinity. They are also employed to prove the measurability of some auxiliary functions. The compactness of W guarantees that for any $\varepsilon > 0$ there exists a finite number of elements of W such that any function in W belongs to the ε -neighborhood of one of these elements of W .

Assumption 2.3 states that convergence with respect to m implies pointwise convergence. This is necessary, together with Assumption 4.1, to prove the continuity of the choice probabilities $P(t|s, z; V, \theta)$ on W . Assumption 2.4 is employed, together with Assumptions 4.1–4.3, to prove the continuity of the choice probabilities in (s, z, θ) .

Assumption 2.5 is necessary to guarantee the identification of (V^*, θ^*) within $(W \times \Theta)$. It is weaker than the commonly made assumption that the subutility V_j^* of each alternative j is independent of the attributes of alternatives other than j . Assumption 2.5 allows the subutility functions of alternatives other than the first to depend on the attributes of all alternatives other than the first. This weakening is important, for example, when preferences incorporate considerations of regret.⁴

Assumption 3.1 is needed to prove the identification of (V^*, θ^*) . Assumption 3.2 is a regularity condition that is used to prove the integrability of several functions in the consistency proof.

Assumptions 4.1–4.3 are used to prove the continuity of the choice probabilities in (s, z, θ) . Assumption 4.1 also guarantees that the probability of ties in (1)

⁴ This example was given by one of the referees.

be zero. In a probit model, for example, Assumption 4.2 imposes restrictions on the covariance matrix of the random terms. Assumption 4.4 implies that the choice probabilities are strictly increasing in the subutility-differences that determine the upper limits of the integral in (1). It also implies that the choice probabilities are strictly positive. Both of these properties are employed in the proof of the identification of (V^*, θ^*) .

Assumptions 5.1–5.3 are made, together with Assumption 2.5, to guarantee the identification of (V^*, θ^*) within $(W \times \Theta)$. The proof of identification also uses the continuity of $P(j|s, z; V, \theta)$ in (s, z, θ) and in the values of V at (s, z) , the strict monotonicity of $P(j|s, z; V, \theta)$ in the utility differences $V_j(s, z) - V_k(s, z)$ ($k = 1, \dots, J, k \neq j$), the strict positivity of $P(j|s, z; V, \theta)$, and Assumption 3.1.

Assumption 5.1 implies that the values of the subutilities of the first alternative are independent of the vector of socioeconomic characteristics, when $z = \bar{z}$. This is analogous to the location normalization, made in linear-in-parameters specifications, which sets to zero all the coefficients of the socioeconomic characteristics in the subutility of the first alternative. The choice of the first alternative as the normalizing alternative is, of course, arbitrary. In the proof of Lemma 8 in the Appendix, we show that Assumptions 5.1 and 2.5 imply that we can recover the function V^* from the choice probabilities $P(1|s, z; V^*, \theta^*), \dots, P(J|s, z; V^*, \theta^*)$. An alternative explanation for this recoverability result is that Assumptions 5.1 and 2.5 allow us to recover the values of the vectors $(V_1^*(s, z), \dots, V_J^*(s, z))$ from the values of the vectors of the subutility differences $(V_2^*(s, z) - V_1^*(s, z), \dots, V_J^*(s, z) - V_1^*(s, z))$. Hence, since as shown in Hotz and Miller (1989), these subutility differences can be recovered from the choice probabilities $P(1|s, z; V^*, \theta^*), \dots, P(J|s, z; V^*, \theta^*)$, Assumption 5.1 and 2.5 allow us to recover the values of the vectors $(V_1^*(s, z), \dots, V_J^*(s, z))$ from the choice probabilities.

Assumption 5.2, which states that the value of all functions in W are equal and known at some point $(\bar{s}, \bar{z}) \in S \times Z$, is made to normalize the scale. This and Assumption 5.3 substitute the standard scale normalization assumptions that fix the values of some coordinates of θ^* . When Assumptions 5.2 and 5.3 are used, it is not necessary to fix these values. Consider, for example, the well known example where $P(1|s, z; V^*, \theta^*) = [1 + \exp[(V_2^*(s, z) - V_1^*(s, z))/\theta^*]]^{-1}$ for $\theta^* \in [\delta, \gamma]$, $\delta, \gamma > 0$. In this case, Assumption 5.3 is satisfied when $\alpha_1 \neq \alpha_2$ since the partial derivative of $P(1|s, z; V^*, \theta)$ with respect to θ when $(s, z) = (\bar{s}, \bar{z})$ is different from zero for all values of $\theta \in [\delta, \gamma]$. Hence, the value of θ^* is identified.

The above assumptions can be imposed, for example, in some multinomial logit, nested logit, multinomial probit, and ordered choice models.

The strong consistency of the estimator is stated in the next theorem, which is proved in the Appendix.

THEOREM 1: *Suppose that the model described in Section 2 satisfies Assumptions 1–5. If, for each $n = 1, 2, \dots, (V_n^{ML}, \theta_n^{ML}) \in (W \times \Theta)$ maximize the likeli-*

hood for n independent observations $\{x^1, \dots, x^n\}$ on the set $(W \times \Theta)$, then

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

4. COMPUTATION OF THE MAXIMUM LIKELIHOOD ESTIMATOR

In this section we present a method to characterize, for each finite number of observations, the set of maximum likelihood estimators of (V^*, θ^*) . The method assumes that the functions in the set W are concave. Additional properties, such as monotonicity, linear homogeneity, and either weak or additive separability can also be incorporated.

The estimator can be computed by transforming the maximization of (2) over $(W \times \Theta)$ into a constrained maximization problem over a Euclidean space.⁵ This transformation is obtained by using the following facts:

(a) *the value of the likelihood function in (2) depends on V only through the values that V attains at the finite number of observed vectors $(s^1, z^1), \dots, (s^n, z^n)$ and*

(b) *the set of values that can be attained by a function in W at a finite number of points can be characterized by a finite number of linear inequalities.*

Fact (a) follows from (1). Fact (b) is formally established in Lemma 1, for a particular set of functions W that satisfies Assumptions 2.5' and 2.6–2.8 below.

ASSUMPTION 2.5': $\forall V \in W, \forall (s, z) \in S \times Z, \forall j, V_j(s, z)$ does not depend on z_k for $k \neq j$.

ASSUMPTION 2.6: $\forall V \in W$ V is concave and monotone increasing.

ASSUMPTION 2.7: $\forall j \in \{1, \dots, J\}$ there exists $B_j = (B_{j,s}, B_{j,z})$ such that $\forall V \in W$ and $\forall (s, z) \in S \times Z$ there exists a subgradient $T_j(s, z)$ of $V_j(s, z)$ such that $-B_j \leq T_j(s, z) \leq B_j$.

ASSUMPTION 2.8: *The set S is bounded below.*

Let W be a set of functions that satisfies Assumptions 2.5', 2.6–2.8, and 5.1–5.2. Let $\{(s^1, z^1), \dots, (s^n, z^n)\}$ be elements of $S \times Z$. Let \underline{s} be such that for all $s \in S$ $\underline{s} \leq s$. Denote α_j by V_j^0 ($j = 1, \dots, J$), \bar{s} by s^0 , \underline{s} by s_1^{n+1} , \bar{z}_j by z_j^0 ($j = 2, \dots, J$), and \bar{z}_1 by z_1^0 and z_1^{n+1} . For all $i = 0, 1, \dots, n, j = 1, \dots, J$, and for $(i, j) = (n + 1, 1)$ let $T_j^i = (T_{j,s}^i, T_{j,z}^i)$ have the same dimensionality as (s^i, z^j) .

LEMMA 1: *The set of all vectors $(V_1^1, \dots, V_J^1; \dots; V_1^n, \dots, V_J^n) \in R^{Jn}$ for which there exists a function V in W such that, for all $i = 1, \dots, n$ and all $j = 1, \dots, J, V_j^i = V_j(s^i, z^i)$ is the set of all vectors $(V_1^1, \dots, V_J^1; \dots; V_1^n, \dots, V_J^n) \in R^{Jn}$ that*

⁵ For calculation methods, see Matzkin (1991).

satisfy the following set of linear inequalities:

$$(L1.1) \quad V_j^i \leq V_j^r + T_{j,s}^r(s^i - s_j^r) + T_{j,z}^r(z_j^i - z_j^r) \quad (i, r = 0, 1, \dots, n; j = 2, \dots, J),$$

$$(L1.2) \quad V_1^i \leq V_1^r + T_{1,s}^r(s^i - s^r) + T_{1,z}^r(z_1^i - z_1^r) \quad (i, r = 0, 1, \dots, n, n + 1)$$

for some $(T_1^0, \dots, T_J^0, T_1^n, \dots, T_J^n, T_1^{n+1})$ satisfying

$$(L1.3) \quad B_j \geq T_j^i \geq 0 \quad (i = 0, 1, \dots, n; j = 1, \dots, J),$$

$$(L1.4) \quad T_{1,s}^{n+1} = 0, \quad \text{and} \quad B_j \geq T_{1,z}^{n+1} \geq 0.$$

The inequalities in (L1.1)–(L1.2) are modifications of the revealed preference conditions, for classical demand data, developed by Afriat (1967a), Diewert (1973), and Varian (1982). Several variations are possible. For example, if the functions in W are assumed to be convex instead of concave in Assumption 2.6, then Lemma 1 will hold with \underline{s} being an upper bound of S instead of a lower bound and with the inequality signs in (L1.1) and (L1.2) reversed. If the functions are not assumed to be monotone increasing, the 0 vectors in the inequalities of (L1.3) and (L1.4) must be substituted, respectively, by $-B_j$ and $-B_{1,z}$. If a k th coordinate ($k > 2$) of the functions in W is assumed to be linearly homogeneous, the following constraints must be added:

$$(L1.5) \quad V_k^i = T_{k,s}^i s^i + T_{k,z}^i z_k^i \quad (i = 0, 1, \dots, n, n + 2),$$

$$(L1.6) \quad s^{n+2} = 0, \quad z_k^{n+2} = 0,$$

and the constraints in (L1.1) corresponding to $j = k$ must be satisfied for $i, r = 0, \dots, n + 1, n + 2$. The equalities in (L1.5) are modifications of the revealed preference conditions developed by Afriat (1972) and Varian (1983). Additional properties can be incorporated in a similar way (see Matzkin (1987)).⁶

By statement (a) and Lemma 1 it follows that when the hypotheses of Lemma 1 are satisfied, the set of estimates of V^* is the set of all functions in W that interpolate between the values $(V_1^{1*}, \dots, V_J^{1*}; \dots; V_1^{n*}, \dots, V_J^{n*})$ and $(T_1^{0*}, \dots, T_J^{0*}, T_1^{n*}, \dots, T_J^{n*}, T_1^{n+1*})$ that solve the maximization of (2) subject to $V_j(s^i, z^i) = V_j^i$ ($i = 1, \dots, n; j = 1, \dots, J$) and the constraints (L1.1)–(L1.4). One such interpolation is given by the function $V = (V_1, \dots, V_J)$ defined by

$$V_1(s, z) = \min \left\{ V_j^{r*} + T_{1,s}^{r*}(s - s_j^r) + T_{1,z}^{r*}(z_j - z_j^r) \mid r = 0, 1, \dots, n, n + 1 \right\}, \quad \text{and}$$

$$V_j(s, z) = \min \left\{ V_j^{r*} + T_{j,s}^{r*}(s - s_j^r) + T_{j,z}^{r*}(z_j - z_j^r) \mid r = 0, 1, \dots, n \right\},$$

for $j = 2, \dots, J$ (see the proof of Lemma 1).

⁶ See also Afriat (1967b, 1973, 1981), Chiappori and Rochet (1987), and Matzkin and Richter (1991) for other revealed preference conditions. Varian (1985) and Epstein and Yatchew (1985) employed revealed preference conditions in statistical tests.

The next Lemma shows that, when Assumption 2.8' holds, the set W in Lemma 1 satisfies the requirements needed for the consistency result.

ASSUMPTION 2.8': $S \times Z$ is compact.

LEMMA 2: Suppose that W satisfies Assumptions 2.2, 2.5', 2.6, 2.7, and 2.8'. Let $m: W \times W \rightarrow R_+$ be the metric determined by the essential supremum norm with respect to $G(\cdot)$, i.e., for all $V, V' \in W$

$$\begin{aligned} m(V, V') &= \text{ess sup } \|V(s, z) - V'(s, z)\| \\ &= \inf \{t | G[\{(s, z) \in S \times Z | \|V(s, z) - V'(s, z)\| > t\}] = 0\}. \end{aligned}$$

Then, W satisfies Assumptions 2.1–2.4 and 5.1–5.2.

5. CONCLUSION

This paper has introduced a strongly consistent semiparametric estimator for polychotomous choice models. The method does not require a parametric structure for the systematic subutility. Instead, it is assumed that this function possesses properties, such as concavity and monotonicity, that are typically assumed in microeconomic theory. The distribution of the unobservable random terms must be specified parametrically.

The estimator is computed by maximizing the likelihood function over a subset of a Euclidean space. The subset is constrained by a finite number of linear inequalities, which are determined by the properties that the nonparametric function is assumed to possess. Some of these properties may reduce considerably the number of variables over which the maximization takes place as well as reduce the number of constraints (see, for example, the concavity constraints in Hildreth (1954) for the case in which the domain of the function lies in the real line).

Although this paper has concentrated on polychotomous choice models, similar estimation techniques can be applied to other limited dependent variable models.

Cowles Foundation for Research in Economics, Department of Economics, Yale University, New Haven, CT 06520-2125, U.S.A.

Manuscript received May, 1987; final revision received September, 1990.

APPENDIX

PROOF OF LEMMA 1: Let V be a function in W . Let z^{n+1} be any vector in Z whose first coordinate is z_1^{n+1} . For all $i = 0, 1, \dots, n, n+1$ and $j = 1, \dots, J$ let T_j^i be a subgradient of V_j at (s^i, z_j^i) such that $-B_j \leq T_j^i \leq B_j$. Then, since V is monotone increasing and satisfies Assumption 5.1, the T_j^i vectors satisfy (L1.3)–(L1.4), and since V is concave and satisfies Assumption 5.2, the vector $\underline{V} = (V_1(s^1, z^1), \dots, V_J(s^1, z^1); \dots; V_1(s^n, z^n), \dots, V_J(s^n, z^n))$ satisfies (L1.1)–(L1.2).

Conversely, suppose that $(V_1^1, \dots, V_J^1; \dots; V_1^n, \dots, V_J^n)$ satisfies (L1.2)–(L1.3), with a vector $(T_1^0, \dots, T_J^0, \dots, T_1^n, \dots, T_J^n, T_1^{n+1})$ that satisfies (L1.3)–(L1.4). Define the functions V_j, V_j ($j = 2, \dots, J$), and V by

$$V_1(s, z) = \min \{V_1^r + T_{1,s}^r(s - s_j^r) + T_{1,z}^r(z_j - z_j^r) | r = 0, 1, \dots, n, n + 1\},$$

$$V_j(s, z) = \min \{V_j^r + T_{j,s}^r(s - s_j^r) + T_{j,z}^r(z_j - z_j^r) | r = 0, 1, \dots, n\},$$

for $j = 2, \dots, J$ and $V(s, z) = (V_1(s, z), \dots, V_J(s, z))$. Then, V satisfies Assumptions 2.5 and 2.7. By (L1.3)–(L1.4) and the definition of V, V is concave and monotone increasing. By the definitions of the V_j functions and (L1.1)–(L1.2), $V(s^i, z^i) = (V_1^i, \dots, V_J^i)$ ($i = 0, 1, \dots, n$) and $V_1(s^{n+1}, \bar{z}) = \alpha_1$. In particular, V satisfies Assumption 5.2. Finally, since $V_1(s^{n+1}, \bar{z}) = \alpha_1$, (L1.3)–(L1.4) and the fact that $s^{n+1} \leq s$ for all $s \in S$ imply, by the definition of $V_1(\cdot)$, that $V_1(s, \bar{z}) = \alpha_1$. Hence, V satisfies Assumption 5.1. Q.E.D.

PROOF OF LEMMA 2: Assumption 2.4 follows from the equicontinuity of the functions in W , which is implied by Assumption 2.7. Assumption 2.3 follows from Assumptions 2.4 and 3.1. By Assumptions 2.3 and 2.7, W is closed with respect to m . Hence, the equicontinuity of the functions in W and Assumptions 5.1 and 5.2 imply by Arzela-Ascoli Theorem (Dunford and Schwartz (1988, pp. 266)) that W is compact with respect to m . Q.E.D.

To prove Theorem 1, let $f(x; V, \theta)$ denote the probability density of x when $(V^*, \theta^*) = (V, \theta)$. Then,

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

The probability measure of $f(x; V^*, \theta^*)$ will be denoted by P^* , the set $\{(d_1, \dots, d_J) | d_j \in \{0, 1\}, \sum_{j=1}^J d_j = 1\}$ will be denoted by D , and the set $D \times S \times Z$ will be denoted by X . Some lemmas are proved next.

LEMMA 3 (Continuity of the Choice Probabilities on $S \times Z$): For all $j \in A$ and all $(V, \theta) \in (W \times \Theta)$, $P(j|s, z; V, \theta)$ is continuous on $S \times Z$.

PROOF: Since, conditional on (s, z) for all $\theta \in \Theta$, $q(\varepsilon; s, z, V, \theta)$ is a Lebesgue density and there exists a Lebesgue integrable function $\phi(\varepsilon)$ such that for all $(s, z, \theta) \in S \times Z \times \Theta$ $|q(\varepsilon; s, z, \theta)| \leq \phi(\varepsilon)$ a.e., it follows by (1), the assumption that any $V \in W$ is continuous on $S \times Z$, and Lebesgue Dominated Convergence Theorem that $P(j|s, z; V, \theta)$ is continuous on $S \times Z$. Q.E.D.

LEMMA 4 (Continuity of Probability Densities on $(W \times \Theta)$): For all $x \in X$, $f(x; V, \theta)$ is continuous on $(W \times \Theta)$.

PROOF: Let (s, z) and $\varepsilon > 0$ be given. Suppose that $\lim_{n \rightarrow \infty} d[(V^n, \theta^n), (V, \theta)] = 0$ where $\{(V^n, \theta^n)\}_{n=1}^\infty \subset (W \times \Theta)$ and $(V, \theta) \in (W \times \Theta)$. Then, from the definition of d and Assumption 2.3, $V_j^n(s, z) \rightarrow V_j(s, z)$ for all $j \in A$. Hence, by (1) and Assumption 4.1 it follows that for all large enough n ,

(L4.1) $|f(x; V^n, \theta) - f(x; V, \theta)| < \varepsilon/2.$

Moreover, by (1), Assumptions 4.1–4.3, Lebesgue Dominated Convergence Theorem, and the hypothesis that $\|\theta^n - \theta\| \rightarrow 0$, it follows that

(L4.2) $|f(x; V^n, \theta^n) - f(x; V^n, \theta)| < \varepsilon/2$

for all large enough n . From (L4.1) and (L4.2) it follows that for all large enough n , $|f(x; V^n, \theta^n) - f(x; V, \theta)| < \varepsilon$. Q.E.D.

LEMMA 5 (Measurability): Define the function $f': (X \times W \times \Theta \times R_{++}) \rightarrow R$ 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: By the definition of d and Assumptions 1.1 and 2.1, there exists a countable dense subset Q of $(W \times \Theta)$ and by Lemma 4

$$f'(x; V, \theta, \varepsilon) = \sup_{(V_n, \theta_n) \in Q} \{f(x; V_n, \theta_n) | d[(V, \theta), (V_n, \theta_n)] < \varepsilon\}.$$

Hence, by the measurability of g , the result follows.

Q.E.D.

LEMMA 6: $\int_X |\log f(x; V^*, \theta^*)| dP^*(x) < \infty$.

PROOF: Immediate from the definition of f , the fact that for all t and all $(s, z) \in S \times Z$ $0 \leq P(t|s, z; V, \theta) \leq 1$, Assumption 3.2, and the fact that the value of $|\log(w)|_w$ over $\{w | K \geq w \geq 0\}$ is bounded when $K < \infty$.

LEMMA 7: Define the function $f^*: (X \times W \times \Theta \times R_{++}) \rightarrow R$ 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 $\varepsilon > 0$

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

is finite.

PROOF: Immediate from the definitions of f' and f^* , Lemma 5, and Assumption 3.2.

LEMMA 8 (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: Suppose that $(V, \theta) \in (W \times \Theta)$ is such that $(V, \theta) \neq (V^*, \theta^*)$. We will show that then there exists $j \in A$ and $(s^*, z^*) \in S \times Z$ such that

$$(L8.1) \quad P(j|s^*, z^*; V, \theta) \neq P(j|s^*, z^*; V^*, \theta^*).$$

We distinguish between three different cases.

Case 1: $\theta \neq \theta^* (V = V^* \text{ or } V \neq V^*)$. In this case (L8.1) follows directly from (1) and Assumptions 5.1–5.3.

Case 2: $\theta = \theta^*$, $V_t = V_t^*$ for all $t \in A$ such that $t \neq 1$, and $V_1 \neq V_1^*$. In this case there exists $(s, z) \in S \times Z$ such that $V_1(s, z) \neq V_1^*(s, z)$ and $V_t(s, z) = V_t^*(s, z)$ for $t = 2, \dots, J$. Denote $V_k(s, z)$ by V_k and $V_k^*(s, z)$ by V_k^* . Then, either $V_1 - V_t > V_1^* - V_t^*$ for all $t > 2$ or $V_1 - V_t < V_1^* - V_t^*$ for all $t > 2$. Since $\theta = \theta^*$ it follows by (1) and Assumption 4.4 that $P(1|s, z; V, \theta) \neq P(1|s, z; V^*, \theta^*)$.

Case 3: $\theta = \theta^*$ and $V_t \neq V_t^*$ for some $t \in A$ such that $t \neq 1$. Let $(s, z) \in S \times Z$ be such that $V_t(s, z) \neq V_t^*(s, z)$. By Assumption 5.1, $V_t(s, \bar{z}) = V_t^*(s, \bar{z})$. Let $z^* \in Z$ be such that $z_1^* = \bar{z}_1$ and $z_t^* = z_t$ ($t \neq 1$). Assumption 2.5 then implies that $V_1(s, z^*) = V_1^*(s, z^*)$ and $V_t(s, z^*) \neq V_t^*(s, z^*)$. For each $k \in A$ denote $V_k(s, z^*)$ by V_k and $V_k^*(s, z^*)$ by V_k^* .

Suppose that $V_t > V_t^*$. (The case in which $V_t < V_t^*$ can be treated in a similar way.) Let $B = \{k \in A | V_k > V_k^*\}$ and $D = \{k \in A | V_k \leq V_k^*\}$. Note that since $t \in B$ and $1 \in D$, $B \neq \emptyset$ and $D \neq \emptyset$. We show that for some $j \in B$ and all $k \in B$

$$V_j - V_k \geq V_j^* - V_k^*.$$

Suppose such j does not exist; then for all k in B there exists k' in B with $V_k - V_{k'} < V_k^* - V_{k'}^*$.

Since the number of elements in B is finite, this implies that there is a cycle k_1, \dots, k_r such that

$$\begin{aligned} V_{k_1} - V_{k_2} &< V'_{k_1} - V'_{k_2}, \\ V_{k_2} - V_{k_3} &< V'_{k_2} - V'_{k_3}, \\ &\dots, \\ V_{k_r} - V_{k_1} &< V'_{k_r} - V'_{k_1}. \end{aligned}$$

Adding up the right-hand side of these inequalities and their left-hand side, we obtain $0 < 0$, which is a contradiction. Hence, since $B \neq \emptyset$, there exists j in B such that for all k in B , $V_j - V_k \geq V_j^* - V_k^*$. Moreover, the definitions of B and D imply that for all k in D ,

$$V_j - V_k > V_j^* - V_k^*.$$

In particular, since $D \neq \emptyset$, $V_j - V_{k'} > V_j^* - V_{k'}^*$ for some $k' \in A$. Since $\theta = \theta^*$ it follows by Assumption 4.4 that $P(j|s^*, z^*; V, \theta) > P(j|s^*, z^*; V^*, \theta^*)$. This completes the proof of (L8.1).

Let now (s^*, z^*) and j satisfy (L8.1). Then, by Lemma 3 there exists a neighborhood N of (s^*, z^*) such that for all $(s, z) \in N$, $P(j|s, z; V, \theta) \neq P(j|s, z; V^*, \theta^*)$. By Assumption 4.4 $P(j|s, z; V^*, \theta^*) > 0$ and by Assumption 3.1 $G(N \cap (S \times Z)) > 0$. Let $E = \{(d, s, z) \in D \times S \times Z | d_j = 1, (s, z) \in N\}$. It then follows that $P^*(E) > 0$ and

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

PROOF OF THEOREM 1: The proof follows by modifying Wald's (1949) result. We sketch the proof.⁷ By Lemmas 6, 7, and 8, for all $(V, \theta) \in (W \times \Theta)$ such that $(V, \theta) \neq (V^*, \theta^*) E \log f(x; V, \theta) < E \log f(x; V^*, \theta^*)$ (Lemma 1 in Wald (1949)); and by Lemmas 4 and 7, for all $(V, \theta) \in (W \times \Theta) \lim_{\varepsilon \rightarrow 0} E \log f'(x; V, \theta, \varepsilon) = E \log f(x; V, \theta)$ (Lemma 2 in Wald (1949)). Hence, for all $(V, \theta) \in (W \times \Theta)$ such that $(V, \theta) \neq (V^*, \theta^*)$, there exists $\varepsilon(V, \theta) > 0$ such that

$$(T1.1) \quad E \log f'(x; V, \theta, \varepsilon(V, \theta)) < E \log f(x; V^*, \theta^*).$$

Let Y be any closed subset of $(W \times \Theta)$ which does not contain (V^*, θ^*) . By Assumptions 1.1 and 2.1, Y is compact. Hence, there exist $\{(V_1, \theta_1), (V_2, \theta_2), \dots, (V_H, \theta_H)\}$ in Y , and positive numbers $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_H$ such that $Y \subseteq \bigcup_{k=1}^H S(V_k, \theta_k, \varepsilon_k)$ and $\varepsilon_k = \varepsilon(V_k, \theta_k)$ where $S(V_k, \theta_k, \varepsilon_k) = \{(v, \theta) \in (W \times \Theta) | d[(V_k, \theta_k), (V, \theta)] < \varepsilon_k\}$. Thus,

$$(T1.2) \quad \sup_{(V, \theta) \in Y} \prod_{i=1}^n f(x^i; V, \theta) \leq \sum_{k=1}^H \prod_{i=1}^n f(x^i; V_k, \theta_k, \varepsilon_k).$$

Hence,

$$(T1.3) \quad \frac{\sup_{(V, \theta) \in Y} \prod_{i=1}^n f(x^i; V, \theta)}{\prod_{i=1}^n f(x^i; V^*, \theta^*)} \leq \frac{\sum_{k=1}^H \prod_{i=1}^n f(x^i; V_k, \theta_k, \varepsilon_k)}{\prod_{i=1}^n f(x^i; V^*, \theta^*)}.$$

By Kolmogorov's Strong Law of Large Numbers and (T1.1), for each $k = 1, \dots, H$,

$$(T1.4) \quad \text{Prob} \left\{ \lim_{n \rightarrow \infty} \sum_{i=1}^n [\log f'(x^i; V_k, \theta_k, \varepsilon_k) - \log f(x^i; V^*, \theta^*)] = -\infty \right\} = 1,$$

⁷ See Kiefer and Wolfowitz (1956) for a similar result.

so that, for each $k = 1, \dots, H$

$$(T1.5) \quad \text{Prob} \left\{ \lim_{n \rightarrow \infty} \frac{\prod_{i=1}^n f^i(x^i; V_k, \theta_k, \varepsilon_k)}{\prod_{i=1}^n f(x^i; V^*, \theta^*)} = 0 \right\} = 1.$$

By (T1.3) and (T1.5) it follows that

$$(T1.6) \quad \text{Prob} \left\{ \lim_{n \rightarrow \infty} \frac{\sup_{(V, \theta) \in Y} \prod_{i=1}^n f(x^i; V, \theta)}{\prod_{i=1}^n f(x^i; V^*, \theta^*)} = 0 \right\} = 1.$$

By Theorem 2 in Wald (1949), this implies that

$$\text{Prob} \left\{ \lim_{n \rightarrow \infty} d[(V_n, \theta_n), (V^*, \theta^*)] = 0 \right\} = 1. \quad Q.E.D.$$

REFERENCES

- AFRIAT, S. (1967a): "The Construction of a Utility Function from Demand Data," *International Economic Review*, 8, 67–77.
- (1967b): "The Construction of Separable Utility Functions from Expenditure Data," mimeo, Purdue University.
- (1972): "The Theory of International Comparisons of Real Income and Prices," in *International Comparison of Prices and Output*, ed. by J. D. Daly. New York: National Bureau of Economic Research.
- (1973): "On a System of Inequalities on Demand Analysis," *International Economic Review*, 14, 460–472.
- (1981): "On the Constructability of Consistent Price Indices between Several Periods Simultaneously," in *Essays in Applied Demand Analysis*, ed. by A. Deaton. Cambridge: Cambridge University Press.
- CHIAPPORI, P., AND J. ROCHET (1987): "Revealed Preference and Differential Demand," *Econometrica*, 55, 687–691.
- COSSLETT, S. R. (1983): "Distribution-free Maximum Likelihood Estimator of the Binary Choice Model," *Econometrica*, 51, 765–782.
- DIEWERT, E. W. (1973): "Afriat and Revealed Preference Theory," *Review of Economic Studies*, 40, 419–426.
- DUNFORD, N., AND J. SCHWARTZ (1988): *Linear Operators, Part I, General Theory*. New York: John Wiley & Sons.
- EPSTEIN, L. G., AND D. J. YATCHEW (1985): "Non-Parametric Hypothesis Testing Procedures and Applications to Demand Analysis," *Journal of Econometrics*, 30, 149–169.
- GALLANT, A. R. (1981): "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form," *Journal of Econometrics*, 15, 211–245.
- (1982): "Unbiased Determination of Production Technologies," *Journal of Econometrics*, 20, 285–323.
- GALLANT, A. R., AND G. H. GOLUB (1984): "Imposing Curvature Restrictions on Flexible Functional Forms," *Journal of Econometrics*, 26, 295–321.
- HAN, A. K. (1987): "Nonparametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimation," *Journal of Econometrics*, 35, 303–316.
- HILDRETH, C. (1954): "Point Estimates of Ordinates of Concave Functions," *Journal of the American Statistical Association*, 49, 598–619.
- HOTZ, V. J., AND R. A. MILLER (1989): "Conditional Choice Probabilities and the Estimation of Dynamic Discrete Choice Models," mimeo, University of Chicago and Carnegie Mellon University.
- ICHIMURA, H. (1988): "Estimation of Single Index Models," University of Minnesota.

- KIEFER J., AND J. WOLFOWITZ (1956): "Consistency of the Maximum Likelihood in the Presence of Infinitely Many Incidental Parameters," *Annals of Mathematical Statistics*, 27, 887–906.
- 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*, 27, 313–334.
- MATZKIN, R. L. (1987): "Semiparametric Estimation of Monotonic and Concave Utility Functions: The Discrete Choice Case," Cowles Foundation Discussion Paper No. 830, Yale University.
- (1990): "Least-concavity and the Distribution-free Estimation of Nonparametric Concave Functions," Cowles Foundation Discussion Paper No. 958, Yale University.
- (1991): "Computation of Concavity Restricted Nonparametric Estimators," mimeo, Cowles Foundation, Yale University.
- (1992): "A Nonparametric and Distribution-free Estimator for the Threshold Crossing and Binary Choice Models," *Econometrica*, forthcoming.
- MATZKIN, R. L., AND M. K. RICHTER (1991): "Testing Strictly Concave Rationality," *Journal of Economic Theory*, 53, 287–303.
- STOKER, T. M. (1986): "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461–1481.
- RAO, PRAKASA (1983): *Nonparametric Functional Estimation*. New York: Academic Press.
- ROBERTSON, T., F. T. WRIGHT, AND R. L. DYKSTRA (1988): *Order Restricted Statistical Inference*. New York: John Wiley and Sons.
- VARIAN, H. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50, 945–973.
- (1983): "Nonparametric Tests of Consumer Behavior," *Review of Economic Studies*, 50, 99–110.
- (1985): "Non-Parametric Analysis of Optimizing Behavior with Measurement Error," *Journal of Economics*, 30, 445–458.
- WALD, A. (1949): "Note on the Consistency of the Maximum Likelihood Estimate," *Annals of Mathematical Statistics*, 20, 595–601.