COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 958

LEAST CONCAVITY AND THE DISTRIBUTION-FREE ESTIMATION OF NONPARAMETRIC CONCAVE FUNCTIONS

Rosa L. Matzkin

October 1990

LEAST CONCAVITY

AND THE DISTRIBUTION-FREE ESTIMATION OF

NONPARAMETRIC CONCAVE FUNCTIONS

by

Rosa L. Matzkin*
Cowles Foundation for Research in Economics
Yale University
New Haven, CT 06520

September 1990

* The support of NSF through grants SES-8720596 and SES-8900291 is gratefully acknowledged. I am grateful to James Heckman, Alvin Klevorick, Quang Vuong and participants in seminars at Bonn, Princeton, Yale, and the 1989 Winter Meeting of the Econometric Society for comments and discussions.

1. INTRODUCTION

Many of the functions studied in economics, such as utility and production functions, are known to be either monotone and concave or monotone and convex. Beyond that characterization, however, there is no specific parametric structure for them. Moreover, in many instances, the values of these functions are only observed through a transformation influenced by some random term. For example, in binary choice models the utility of each alternative consists of a systematic part, which is a function of observable attributes, and a random part. Both parts of the utility function are unobservable. One can only observe whether the utility difference between the two alternatives is positive. In certain situations, economic theory can provide some indications about the monotonicity and concavity of the systematic part. If the distribution of the random part does not belong to a parametric family, however, this knowledge will typically not be sufficient to identify the systematic part of the function.

This lack of identification contrasts with the case in which either or both the systematic function of observable variables and the distribution of the random part belong to parametric families. Matzkin (1987) showed that when the distribution of the random term belongs to a parametric family, it is possible to identify and consistently estimate the systematic function of exogenous variables, which is assumed to be nonparametric, monotone, and concave. Manski (1975,1985), Cosslett (1983), Stoker (1985), Han (1987), Ichimura (1986), and Klein and Spady (1986), among others, showed that when the systematic function of exogenous variables belongs to certain parametric families, it is possible to identify and consistently estimate the parameters of the systematic function, even when the distribution of the random terms is not assumed to have any particular parametric structure.

In this paper, I analyze the identification and estimation of three microeconometric models in which neither the systematic function of observable exogenous variables nor the distribution of the random term belongs to a parametric family of functions or distributions. The main result is that when the systematic function is a monotone and concave function whose domain is a compact set, it is possible, under certain weak additional conditions, to identify and estimate a representative of the systematic function. This representative is a monotone transformation of the systematic function. Hence, its isovalue sets coincide with the isovalue sets of the systematic function.

The result of this paper complements the work of Matzkin (1988), where it was shown that when the systematic function possesses properties such as homogeneity of degree one or certain types of separability, it is possible to identify and consistently estimate both the systematic function and the distribution of the random term in binary threshold and choice models. (See also Matzkin (1990a)). When the systematic function is monotone and concave, estimation of this function by the methods developed in Matzkin (1988) requires strong restrictions on the shape of its isovalue sets. In contrast, the method developed in this paper only imposes minimal restrictions on the isovalue sets of the systematic function.

The three microeconometric models studied are a binary choice model, a generalized regression model, and a threshold crossing model. The random term in the binary choice model is assumed to be median independent of the exogenous observable variables, while the random terms in the generalized regression model and the threshold crossing model are independent of the observable exogenous variables. The distributions of the random terms are otherwise unknown. The systematic functions are monotone increasing and concave; they are also required to posses other weak properties, which will be made explicit in the following sections. In the binary choice and generalized regression models, I study the estimation of a representative of the systematic function, while in the threshold crossing model, I study the estimation of representatives of both the systematic function and the distribution of the random term.

The approach proceeds by characterizing each subset of observationally equivalent systematic functions by a unique least concave representative in a compact set of least concave representatives. This particular characterization has three appealing properties. First, the isovalue sets of the systematic functions can be obtained from the isovalue sets of their least concave representatives. Second, the compactness of the set of least concave functions allows us to devise consistent estimators for the representative of the systematic function. Third, least concave representatives enable us easily to obtain a consistent estimator of the equivalence class to which the systematic function belongs from a consistent estimator of its least concave representative. Being able to estimate the equivalence class can be important when it is expected that information that will be available later will help reduce the equivalence class.

Least concave functions have been previously studied in the theory of representations of concavifiable preference relations (deFinneti (1949), Debreu (1976), Kannai (1977,1980)). When a preference relation can be represented by a concave utility function, one typically asks whether a least concave utility function for that preference exists. A function h is a least concave utility function for a preference relation if any concave utility function for the same preference is a concave and strictly increasing transformation of h. Hence, the entire set of concave utility functions for a given preference relation can be obtained by composing the least concave utility function for that preference with any possible concave and strictly increasing function. Debreu (1976) showed that any preference relation that admits a concave utility function admits a least concave utility function.

I show that a least concave representative of the systematic function generating the observations is identified within a set of least concave functions. I develop distribution-free methods of consistently estimating this least concave representative.

To make the new estimation methods operational, a technique for calculating the estimators is developed. The technique proceeds by first maximizing a criterion function over a set of concave functions and second obtaining the values of a least concave representative that generates, on the observed vectors, the same preorder generated by this concave function. The first step adapts methods studied in Matzkin (1987). The second step involves solving a linear programming problem. By using this technique we can approximate the values, at each observed vector of exogenous variables, and the gradients, at every point and up to a positive constant, of a least concave representative that maximizes the criterion function.

The outline of the paper is as follows. In the next section, I present a binary choice model and show that in that model the systematic subutility function can not be identified. In Section 3, I present a set of least concave functions and show that sets of observationally equivalent functions can be characterized by a unique representative in that set. A method of estimating the least concave representative of the systematic function is presented in Section 4 and the calculation of this estimator is studied in Section 5. Sections 6 and 7 show that a similar estimation approach can be applied to generalized regression models and binary threshold crossing models. The main conclusions of the paper are summarized in Section 8. All the proofs are presented in the Appendix.

2. A NONIDENTIFIED MODEL

In this section, I present an example of a microeconometric model in which it is not possible to identify the systematic function of the observable exogenous variables. The example is a fully nonparametric binary choice model with a median independent random term.

Binary choice models have been employed to analyze a wide variety of problems in which an economic agent has to choose between two alternatives. Examples include the choice between two modes of transportation and the choice between purchasing or not purchasing a commodity.

In these models, the value of an observable dichotomous variable y is determined by the value of an unobservable latent variable y^* according to

$$y = \mathrm{sgn}(y^*)$$

where $\mathrm{sgn}(y^*) = 1$ if $y^* \geq 0$ and $\mathrm{sgn}(y^*) = -1$ otherwise. The value of the variable y^* depends on a pair of vectors of observable exogenous variables (x_1, x_2) and an unobservable random variable η through the relationship:

$$y^* = h^*(x_1) - h^*(x_2) - \eta .$$

The pair $x = (x_1, x_2)$ possesses a marginal probability measure P_x whose support will be denoted by $X \times X$.

In a choice of transportation example, x_1 and x_2 denote, respectively, the observable attributes of the first and second mode; h^* denotes the utility of the observable attributes; y^* denotes the difference between the utilities of the two alternative modes; and η is the difference between the unobservable random subutilities of the modes.

I consider a particular binary choice model in which it is assumed that $h^*: X \to R$ is monotone and concave but otherwise unknown and the marginal distributions $F^*_{\eta|x}$ of η are unknown but, for all $x = (x_1, x_2)$, satisfy $F^*_{\eta|x}(0) = .5$.

This weak set of assumptions is not restrictive enough to identify h^* from observations on (y, x_1, x_2). To show this, we let $P(j|x_1, x_2; h^*, F^*)$ denote the probability of choosing alternative j (j=1,2) given the vector of attributes (x_1, x_2); $P(j|x_1, x_2; h^*, F^*) = F^*_{\eta|x}(h^*(x_1) - h^*(x_2))$ (j=1,2; $x=(x_1, x_2) \in X{\times}X$). The most information we can obtain from observations on (y, x_1, x_2) is the function $P(j|x_1, x_2; h^*, F^*)$. A concave and monotone function $h: X \to R$ is observationally equivalent to h^* if, given our knowledge about the properties of h^* and $F^*_{\eta|x}$, it is not possible to distinguish h^* from h. This will be true if we can find a distribution $F_{\eta|x}$ that has the same properties that $F^*_{\eta|x}$ is known to possess and is such that for every $x \in (X{\times}X)$, $P(j|x_1, x_2; h^*, F^*) = P(j|x_1, x_2; h, F)$ for j=1,2. Formally, DEFINITION 1: *A concave and monotone function* $h: X \to R$ *is* <u>*observationally equivalent*</u> *to* h^* *if there exists a distribution function* $F_{\eta|x}$ *satisfying* $F_{\eta|x}(0) = .5$ *for every* $x = (x_1,x_2) \in (X \times X)$ *and such that*

$$P(1|x_1,x_2;h^*,F^*) = P(1|x_1,x_2;h,F).$$

DEFINITION 2: *The function* h^* *is* <u>*identified*</u> *within a set* $\tilde{W}$ *of concave and monotone functions* $h: X \to R$ *if* $h^* \in \tilde{W}$ *and there is no* $h \in \tilde{W}$ *such that* $h \neq h^*$ *and* h *is observationally equivalent to* h^*.

The next lemma states that h^* is not identified within the set of concave and monotone functions defined on X.

LEMMA 1: *There exists a concave and monotone function* $h: X \to R$ *that is observationally equivalent to* h^* *and is such that* $h \neq h^*$.

3. EQUIVALENCE CLASSES AND LEAST CONCAVE REPRESENTATIVES

One possible way to deal with the nonidentification of h^* is to look for additional restrictions on h^* that guarantee that h^* is identified within the smaller set of functions that satisfy those restrictions. This was the approach taken in Matzkin (1988). In this paper, I propose to deal with the nonidentification of h^* in a different way.

The method proposed in this paper studies a least concave representative of the class of functions that are observationally equivalent to h^*. This representative can be strongly consistently estimated. From this estimator we can obtain estimators for the isovalue sets of h^* and for the equivalence class of h^*.

In the next subsection I present the definition of least concave functions and study some of their properties.

3.1. LEAST CONCAVE FUNCTIONS

To explain the concept of a least concave function I follow Debreu (1976): Let $h:X \to R$ be a concave function on a convex set $X \subset R^K$ and consider the set U of all monotone transformations of h. Define on U the relation "u is more concave than v" by "there exists a concave function $g:R \to R$ such that $u = g \cdot v$."[2] A least concave representative of U is a least element of U with respect to the relation "more concave than." The next definition states this formally.

DEFINITION 3: *Suppose that $h:X \to R$ is a concave function, where $X \subset R^K$ is a convex set. A function $\bar{h}:X \to R$ is a* <u>*least concave*</u> *representative of h if for any strictly increasing function $f:R \to R$ such that $f \cdot h$ is concave there exists a concave function $g:R \to R$ such that*

$f \cdot h = g \cdot \bar{h}$.

In other words, $\bar{h}$ is a least concave representative of h if any concave function that is a monotone transformation of h is a concave transformation of $\bar{h}$. Note that the function $g(\cdot)$ in Definition 3 is necessarily strictly increasing. As an example, suppose that $h:R^2 \to R$ is given by the strictly concave function $h(x_1,x_2) = \log(x_1) + \log(x_2)$. Then, a least concave representative of h is given by the linearly homogeneous function $\bar{h}(x_1,x_2) = (x_1 \cdot x_2)^{1/2}$. (See Kannai (1980).)

Note that all functions that are monotone transformations of h possess the same isovalue sets as h. In particular, since h is a monotone transformation of its least concave representative $\bar{h}$, the isovalue sets of h coincide with the isovalue sets of $\bar{h}$, i.e. for all $\tilde{x} \in X$,

$$\{ x \in X \mid h(x) = h(\tilde{x}) \} = \{ x \in X \mid \bar{h}(x) = \bar{h}(\tilde{x}) \}.$$

Definition 3 immediately implies that any two concave functions are strictly increasing transformation of each other if and only if they possess the same least concave representatives.

LEMMA 2: Let $h:X \to R$ and $h':X \to R$ be concave and let $\bar{h}$ be a least concave representative of h . There exists a strictly increasing function $f:R \to R$ such that $h' = f \circ h$ if and only if $\bar{h}$ is a least concave representative of h'.

The set U of concave functions that are monotone transformations of h is not complete with respect to the relation "more concave than"; i.e., given any two functions h,h' in U, it is possible that h is not more concave than h' and h' is not more concave than h. Nevertheless, U possesses a least element with respect to this relation. This surprising result, which is due to Debreu (1976), is stated in the next lemma.

LEMMA 3 (Existence): *Suppose that $h:X \to R$ is a concave function on a convex set $X \subset R^K$. Then, there exists a least concave representative $\bar{h}$ of h.*

Least concave representations are unique up to increasing linear transformations. This fact, which can be easily shown, is stated as a lemma for further reference.

LEMMA 4 (Uniqueness): *Suppose that* $h:X \rightarrow R$ *is a concave function on a convex set* $X \subset R^K$. *If* $\bar{h}$ *and* $\bar{h}'$ *are both least concave representatives of* h, *then there exist* $a \in R$ *and* $b \in R_{++}$ *such that for all* $x \in X$

$$\bar{h}(x) = a + b\,\bar{h}'(x) \ .$$

The next two lemmas show that least concave representatives possess certain properties of the functions they represent.

LEMMA 5 (Linearity): *Suppose that* $h:X \rightarrow R$ *is linear and* $\bar{h}:X \rightarrow R$ *is a least concave representative of* h. *Then* $\bar{h}$ *is linear.*

LEMMA 6 (Continuous Differentiability): *Suppose that* $h:X \rightarrow R$ *is a concave function,* $\bar{h}$ *is a least concave representative of* h, *and* $x \in int(X)$ *is such that* $h(x) < h(x')$ *for some* $x' \in X$. *If* h *is continuously differentiable at* x *then* $\bar{h}$ *is continuously differentiable at* x.

I next define a compact set of C^2 least concave functions. The compactness of this set will be employed to obtain a nonparametric estimator for the least concave representative of h^*.

To guarantee the compactness of the set of functions, we need to impose some uniform bounds on the values, gradients, Hessians, and Gaussian curvatures of the functions in the set and we need to require some Lipschitz properties of the functions and their gradients and Hessians.

Let $\underline{x}$ and $\bar{x}$ be elements of R^K and let X denote the set $\{ x \in R^K \mid \underline{x} \leq x \leq \bar{x} \}$.

For any C^2 concave function $h:X \to R$, let $Dh(x)$ and $D^2h(x)$ denote, respectively, the gradient and Hessian of h at $x \in X$, let $D_j h(x)$ and $D^2_{ij}h(x)$ denote respectively the jth and ijth elements of $Dh(x)$ and $D^2h(x)$, and let $c(h,x)$ denote the Gaussian curvature[1] of h at x. The functions h, $Dh(x)$, and $D^2h(x)$ are called V-Lipschitzian ($V > 0$) if for all $x,y \in X$, and $i,j=1,\ldots,K$, $|h(x) - h(y)| \leq V \|x - y\|$, $|D_j h(x) - D_j h(y)| \leq V \|x - y\|$, and $|D^2_{ij}h(x) - D^2_{ij}h(y)| \leq V \|x - y\|$, respectively.

To impose bounds on the values, subgradients, Hessians, and Gaussian curvatures of the functions, let α, γ, and c be strictly positive real numbers with $\alpha < \gamma$ and $c > 0$. Let V_2 and V_3 be strictly positive numbers. Let $B^L, B^U \in R^K$ be such that $0 \leq B^L_k < B^U_k$ $(k=1,\ldots,K)$ and $B^L_K > 0$. Let $A = (A_{ij})$ and $C = (C_{ij})$ be $K \times K$ matrices such that $A_{ij} \leq C_{ij}$ $(i,j=1,\ldots,K)$. Require, further, that A, C, B^L, B^U, and c satisfy that for all matrices $T = (T_{ij})$ and vectors t such that $A_{ij} \leq T_{ij} \leq C_{ij}$ $(i,j=1,\ldots,K)$ and $B^L_j \leq t_j \leq B^U_j$ $(j=1,\ldots,J)$, the determinant of the matrix

$$\begin{pmatrix} -T & t \\ -t^T & 0 \end{pmatrix} \frac{1}{\|t\|^{K+1}}$$

is not smaller than c.

We can now employ this to define a compact set W. We let W be the set of all functions $h: X \to R$ such that

(i) h is a C^2 least concave representative of some concave function $\hat{h}: X \to R$,

(ii) $h(\underline{x}) = \alpha$, $h(\bar{x}) = \gamma$,

(iii) $\forall x \in X \ \forall i,j \quad B_j^L \le D_j h(x) \le B_j^U$, $A_{ij} \le D_{ij}^2 h(x) \le C_{ij}$, and

(iv) $Dh(\cdot)$ is V_2-Lipschitzian and $D^2h(\cdot)$ is V_3-Lipschitzian.

The set of concave functions that possess least concave representatives in W will be denoted by $\tilde{W}$. Functions in $\tilde{W}$ are monotone increasing, concave, and strictly increasing with respect to the Kth coordinate. By Lemma 4 and condition (ii), all functions in $\tilde{W}$ possess a unique representative in W.

Let d denote the C^2 distance, which is defined by

$$\begin{aligned} d(h,h') = \ & \sup_{x \in X} \ \| h(x) - h'(x) \| \\ & + \sup_{x \in X, j=1,\ldots,K} \ \| D_j h(x) - D_j h'(x) \| \\ & + \sup_{x \in X, i,j=1,\ldots,K} \ \| D_{ij}^2 h(x) - D_{ij}^2 h'(x) \| . \end{aligned}$$

The next theorem establishes the compactness of W with respect to d.

THEOREM 1: *W is compact with respect to the metric d.*

3.2. CHARACTERIZATION OF EQUIVALENCE CLASSES USING LEAST CONCAVE REPRESENTATIVES

The set of least concave functions presented above can be employed to characterize sets of observationally equivalent concave functions. I next show this for the binary choice model that was presented in Section 2. In Sections 6 and 7, I apply the same characterization to a generalized regression model and a binary threshold crossing model.

The next theorem shows that if $h^* \in \tilde{W}$, then the set of functions in $\tilde{W}$ that are observationally equivalent to h^* possess a common (and unique) least concave representative in W.

THEOREM 2: *Suppose that* $h, h^* \in \tilde{W}$. *Let* $\bar{h}$ *and* $\bar{h}^*$ *denote respectively the least concave representatives in* W *of* h *and* h^*. *Then,* h *is observationally equivalent to* h^* *if and only if* $\bar{h} = \bar{h}^*$.

Hence, the set of functions that are observationally equivalent to h^* can be characterized by the least concave representative of h^* in W. From this result and Lemma 2 it follows that h is observationally equivalent to h^* if and only if h is a strictly increasing transformation of h^*. Note that the isovalue sets of all functions in $\tilde{W}$ that are observationally equivalent to h^* coincide with the isovalue sets of the least concave representative $\bar{h}^*$ of h^*. From Lemma 4 and condition (ii) in the definition of W, the least concave representative in W of h^* is unique.

The definition of least-concavity together with Theorem 2 and the definition of $\tilde{W}$ imply that

$$\{ h \in \tilde{W} \mid h \text{ is observationally equivalent to } h^* \}$$

$$= \{ h: X \to R \mid h = g \cdot \bar{h}^* \text{ for some concave and strictly increasing function } g \}.$$

Hence, if we obtain an estimator for the least concave representative, $\bar{h}^*$, of h^* we can easily obtain an estimator for the equivalence class to which h^* belongs. In the next section, I show how to obtain an estimator for $\bar{h}^*$.

4. CONSISTENT ESTIMATION OF A LEAST CONCAVE REPRESENTATIVE OF h^*

To obtain an estimator for $\bar{h}^*$, we can modify the Maximum Score Estimation Method, which was introduced by Manski (1975, 1985). Manski developed this method to estimate a parametric function h^* in a binary choice model with median independent random term. In the model that Manski studied, h^* was assumed to be a linear function and F^* was assumed to be unknown.

Following Manski (1975, 1985), we define the *population score function* by

(1) $S(h)$

$$= E[\ y\ \text{sgn}(h(x_1) - h(x_2))\ ]$$

$$= P[\ y^* \geq 0,\ \ h(x_1) - h(x_2) \geq 0\ ] + P[\ y^* < 0,\ \ h(x_1) - h(x_2) < 0\ ]$$

$$- P[\ y^* \geq 0,\ \ h(x_1) - h(x_2) < 0\ ] - P[\ y^* < 0,\ \ h(x_1) - h(x_2) \geq 0\ ].$$

and we define the *sample score function* by

(2) $S_N(z^{(N)}, h)$

$$= N^{-1} \Sigma_{i=1}^{N}\ y^i\ \ \text{sgn}(\ h(x_1^i) - h(x_2^i)\ )$$

$$= P_N[\ y^* \geq 0,\ \ h(x_1^i) - h(x_2^i) \geq 0\ ]\ +\ P_N[\ y^* < 0,\ \ h(x_1^i) - h(x_2^i) < 0\ ]$$

$$- P_N[\ y^* \geq 0,\ \ h(x_1^i) - h(x_2^i) < 0\ ]\ -\ P_N[\ y^* < 0,\ \ h(x_1^i) - h(x_2^i) \geq 0\ ],$$

where P_N denotes the empirical distribution of the vector (y,x_1,x_2) and $\{\ (y^i,x_1^i,x_2^i)\ \}_{i=1}^N$ are observations on the vector (y,x_1,x_2).

We define the *least concave maximum score* estimator to be the function $\hat{h}_N\colon X \to R$ that maximizes $S_N(z^{(N)},\cdot)$ over the set W.

Note that from (2) it follows that the value that $S_N(z^{(N)},\ \cdot)$ attains at any function h depends only on the preordering induced by that function on $x^1,\ldots,x^N$. That is, any two functions h and h' that are such that for any i,j $[\ h(x_i) \geq h(x_j)\ \ \Leftrightarrow\ \ h'(x_i) \geq h(x_j)\ ]$ will yield the same value of S_N. This property will be employed in the computation of $\hat{h}_N$.

That the estimator $\hat{h}_N$ is strongly consistent will be shown under the following set of assumptions:

ASSUMPTION A.1: For all $(x_1,x_2)\in(X\times X)$ $\text{MEDIAN}[y^*|x_1,x_2] = h^*(x_1) - h^*(x_2)$.

ASSUMPTION A.2: For all $(x_1,x_2) \in (X\times X)$ $0 < \Pr(\ y^* \geq 0\ |\ x_1,x_2) < 1$.

ASSUMPTION A.3: The observable vector $x = (x_1,x_2)$ possesses an absolutely continuous probability measure P_x, whose support is $(X\times X)$.

ASSUMPTION A.4: The observations $\{\ (y^i,x_1^i,x_2^i)\ \}_{i=1}^N$ are independent.

ASSUMPTION A.5: $h^* \in \tilde{W}$.

These assumptions guarantee that $\bar{h}^*$ is identified within W (Theorem 3) and $\hat{h}_N$ is a consistent estimator of $\bar{h}^*$ (Theorem 4).

THEOREM 3 (Identification): *Suppose that Assumptions A.1-A.5 are satisfied. Then,* $\bar{h}*$ *uniquely maximizes the population score function* $S(\cdot)$ *over* W.

THEOREM 4 (Consistency): *Suppose that Assumptions A.1-A.5 are satisfied. Then, with probability one,* $lim_{N\to\infty} d(\hat{h}_N, \bar{h}^*) = 0$.

To calculate $\hat{h}_N$, we need to maximize S_N over W. In the next section I present a technique to solve this maximization problem.

5. MAXIMIZATION OF CRITERION FUNCTIONS OVER SETS OF LEAST CONCAVE FUNCTIONS

In this section I describe a technique to find a least concave function that maximizes the value of a criterion function. I consider criterion functions whose values at any concave function h depend only on the preordering that is induced by h on a finite number of observed points on X. The criterion function S_N considered in the above section, for example, satisfies this property.

DEFINITION 4: *Let* $\{x^1,\ldots x^N\}$ *denote the set of observed points in* X. *Let* $z^{(N)} = \{ (y^i,x^i) \}_{i=1}^N$. *A criterion function* $S_N(z^{(N)},\cdot):\hat{W} \to R$ *will be said to be* <u>order-dependent</u> if $S_N(z^{(N)}, h) = S_N(z^{(N)}, h')$ whenever for all $i,j \in \{1,\ldots,N\}$ $[\ h(x^i) \geq h(x^j) \iff h'(x^i) \geq h'(x^j)\ ]$

In other words, a criterion function $S_N(z^{(N)}, \cdot)$ is order-dependent if it attains the same value at any two functions that generate identical preorderings on $\{ x^1,\ldots,x^N \}$.

When a criterion function $S_N(z^{(N)},\cdot)$ is order-dependent, it is possible to divide the problem of maximizing $S_N(z^{(N)},\cdot)$ over any set $\bar{W}$ of least concave functions into two separate problems. In the first problem, $S_N(z^{(N)},\cdot)$ is maximized over a set of concave functions, $\tilde{W}$, whose least concave representatives characterize the set $\bar{W}$. In the second problem, a least concave function $\bar{h}$ in $\bar{W}$ is found, that induces the same preordering on the vectors $\{x^1,\ldots,x^N\}$ as the concave function found in the first problem. Since $S_N(z^{(N)},\cdot)$ is order-dependent and $\bar{h}$ induces the preorder that was found to maximize $S_N(z^{(N)},\cdot)$, it follows that $\bar{h}$ maximizes $S_N(z^{(N)},\cdot)$.

The next theorem shows how to find the values and subgradients[3] of a least concave function that induces on $\{ \underline{x},x^1,\ldots,x^N,\bar{x} \}$ the same preordering as a function $g(\cdot)$ does. The function $g(\cdot)$ is assumed to be concave, monotone increasing, and strictly increasing in the Kth coordinate, and the least concave function is restricted to attain the values α and γ, respectively, at $\underline{x}$ and $\bar{x}$.

THEOREM 5: *Let* $g(\underline{x}),g(x^1),\ldots,g(x^N),g(\bar{x})$ *be the values of a function at* $\underline{x},x^1,\ldots,x^N,\bar{x}$. *Assume that* g *is monotone increasing, concave, and strictly increasing with respect to the Kth coordinate. Let* $p \in R^N_{++}$ *and* *consider the following minimization problem:*

$$\text{Minimize} \quad (3) \quad \sum_{i=1}^{N} p_i \ \bar{h}_i$$

subject to

$$(4) \quad \bar{h}_i \leq \bar{h}_j + \bar{\beta}_j (x^i - x^j) \qquad i,j=0,\ldots,N+1,$$

$$(5) \quad \bar{\beta}_j \geq 0, \quad \bar{\beta}_{j,K} > 0 \qquad i,j=0,\ldots,N+1,$$

$$(6) \quad \bar{h}_i < \bar{h}_j \quad \text{if} \quad g(x^i) < g(x^j) \qquad i,j=0,1,\ldots,N,N+1,$$

$$(7) \quad \bar{h}_i = \bar{h}_j \quad \text{if} \quad g(x^i) = g(x^j) \qquad i,j=0,1,\ldots,N,N+1,$$

$$(8) \quad x^0 = \underline{x}, \quad x^{N+1} = \bar{x}, \quad \bar{h}^0 = \alpha, \quad \text{and} \quad \bar{h}_{N+1} = \gamma$$

Let $h^*_0, h^*_1, \ldots, h^*_N, h^*_{N+1}$ *and* $\beta^*_0, \ldots, \beta^*_{N+1}$ *be a solution to this minimization problem. Then, there exists a least concave function* $\bar{h}$ *such that*

(i) $\bar{h}$ *generates on* $\{\underline{x}, x^1, \ldots, x^N, x^{N+1}\}$ *the same preordering as* g *does,*

(ii) $\bar{h}(x_i) = h^*_i \quad (i=0,1,\ldots,N,N+1)$,

(iii) for a.e. $x \in X$, *the subgradient* $D\bar{h}(x)$ *of* $\bar{h}$ *at* x *satisfies* $D\bar{h}(x) = a(x)\ \beta^*_j$ *, where* $a(x) > 0$ *is a real number and* j *is such that* $h^*_j + \beta^*_j (x - x^j) = \min \{ h^*_i + \beta^*_i (x - x^i) \mid i=0,1,\ldots,N+1 \}$.

Theorem 5 describes how to obtain the values at each x^i and the gradients, at a.e. x, of a least concave representative. We note that the choice of p is irrelevant as long as $p \in R^N_{++}$ and that, typically, many of the constraints in (6)-(8) will be redundant, so they can be eliminated.

The above result suggests a way of obtaining a least concave function that maximizes $S_N(z^{(N)},\cdot)$ over W. First find the values of a function $g(\cdot)$ that maximizes $S_N(z^{(N)},\cdot)$ over the set $\tilde{W}$. Second, use the values $g(\underline{x}),g(x^1),\ldots,g(x^N),g(\bar{x})$ to obtain the values of a least concave function that maximizes $S_N(z^{(N)},\cdot)$ over W, by using the result of Theorem 5.

Unfortunately, the full characterization of the set $\tilde{W}$ is unknown, because the functions in W are required to satisfy conditions (iii)-(iv) in the definition of W (see Section 3). Hence, instead of maximizing $S_N(z^{(N)},\cdot)$ over $\tilde{W}$, we propose to maximize $S_N(z^{(N)},\cdot)$ over the larger set of concave and monotone functions that are strictly increasing in their Kth argument. The solution to the second step obtained from the maximizer function g of such first step maximization may not belong to W. If such is the case, a new function g yielding perhaps a lower value of $S_N(z^{(N)},\cdot)$ must be selected till a solution of the second step maximization is found.

To find a function $g(\cdot)$ that maximizes $S_N(z^{(N)},\cdot)$ over a set of concave and monotone functions that are strictly increasing in their Kth argument, we can follow the technique presented in Matzkin (1987). We find real valued $\hat{h}^0,\ldots,\hat{h}^N,\hat{h}^{N+1}$ and vector valued $D\hat{h}^0,\ldots,D\hat{h}^N,D\hat{h}^{N+1}$ that solve the following problem:

maximize (9) $$S_N(\, z^{(N)},\hat{h}^0,\ldots,\hat{h}^N,\hat{h}^{N+1}\,)$$

subject to

(10) $$\hat{h}^i \le \hat{h}^j + D\hat{h}^j\,(x^i - x^j) \qquad i,j=0,1,\ldots,N,N+1,$$

(11) $$0 \le D\hat{h}^i, \quad 0 < D\hat{h}^i_K \qquad i=0,1,\ldots,N,N+1.$$

The algorithm developed in Matzkin (1990b) can be used to calculate a solution to this problem. This algorithm looks for points inside the constraint set that yield larger values of the objective function by searching along randomly chosen segments whose endpoints are determined by parametric concave functions.

To find the values of a least concave function that belongs to W and induces on $\{ \underline{x}, x^1, \ldots, x^N, \bar{x} \}$ the same preordering as the function g does, we can solve the following problem, which is based upon the result in Theorem 5:

$$\text{Minimize} \quad (12) \quad \sum_{i=1}^{N} p_i \; \hbar_i$$

subject to

$$(13) \quad \hbar_i < \hbar_j \quad \text{if} \quad g(x^i) < g(x^j) \qquad i,j=0,\ldots,N+1,$$

$$(14) \quad \hbar_i = \hbar_j \quad \text{if} \quad g(x^i) = g(x^j) \qquad i,j=0,\ldots,N+1,$$

$$(15) \quad \hbar^i \leq \hbar^j + D\hbar^j \, (x^i - x^j) \qquad i,j=0,1,\ldots,N,N+1,$$

$$(16) \quad B^L \leq D\hbar^j \leq B^U \qquad i=0,1,\ldots,N,N+1,$$

$$(17) \quad x^0 = \underline{x}, \quad x^{N+1} = \bar{x}, \quad \hbar^0 = \alpha, \quad \text{and} \quad \hbar_{N+1} = \gamma$$

$$(18) \quad A \leq D^2\hbar^{i,j} \leq C \qquad i,j=0,1,\ldots,N,N+1,$$

$$(19) \quad | \, D\hbar^j_k - D\hbar^i_k \, | \leq V_2 \, \| x^j - x^i \| \qquad i,j=0,\ldots,N+1; k=1,\ldots,K,$$

$$(20) \quad | \, D^2\hbar^{jt}_{k,r} - D^2\hbar^{is}_{k,r} \, | \leq V_3 \, \| x^j - x^i \| \qquad k=1,\ldots,K,, \; i,j,t,s=0,1,\ldots,N+1,$$

where $D^2\hbar^{i,j}$ is a matrix whose k,r element is

$$D^2\hat{h}^{ij}_{k,r} = (D\hat{h}^j_k - D\hat{h}^i_k) \;/\; (x^j_r - x^i_r).$$

The vector p can be chosen arbitrarily from R^K_{++}. The values of $\hat{h}^i$, $D\hat{h}^i$, and $D^2\hat{h}^{ij}$ are interpreted as the values, gradients, and Hessians of a function $\hat{h}$. The constraints in (15) restrict $\hat{h}^0,\ldots,\hat{h}^N,\hat{h}^{N+1}$ and $D\hat{h}^0,\ldots,D\hat{h}^N,D\hat{h}^{N+1}$ to be, respectively, the values and gradients of a concave function. The constraints in (16) impose the monotonicity of $\hat{h}$ and the uniform boundedness of its gradients. Constraint (17) guarantees that $\hat{h}$ satisfies condition (ii) in the definition of W. The concavity of $\hat{h}$ together with (17) and the uniform bounds on the subgradients guarantee the uniform Lipschitz property on the values of the function. Constraint (18) is necessary for the uniform boundedness of the Hessians of $\hat{h}$, and constraints (19) and (20) are necessary for the Lipschitzian restrictions on the gradients and Hessians of $\hat{h}$. It is possible that a function $\hat{h}$ satisfying constraints (18)-(20) does not satisfy at every point of its domain conditions (iii)-(iv) in the definition of W. This problem tends to disappear, however, when the number of observations is large.

The solution to the above problem provides us then with an approximation of the values at $\underline{x},x^1,\ldots,x^N,x^{N+1}$ and the subgradients, at a.e. point and up to a positive constant, of a least concave function that maximizes $S_N(z^{(N)},\cdot)$ over W.

As a final comment in this section, we note that constraints (16)-(20) could be added to the maximization problem described by (9)-(11), without reducing the set of least concave functions over which $S_N(z^{(N)},\cdot)$ is maximized. Since W is included in the set of concave and monotone functions satisfying conditions (ii)-(iv) in the definition of W, the set of least concave representatives of the functions in this set necessarily includes W. Imposing constraints (16)-(20) in the maximization problem described by (9)-(11) will, in particular, be useful when the solution to the second step maximization obtained from the solution of (9)-(11) does not exist.

6. A GENERALIZED REGRESSION MODEL

The nonidentification of the systematic function h^* that was shown to hold for the binary choice model in Section 2 can occur in various other models. I present two such other models, a generalized regression model and a binary threshold crossing model. In these models, neither the systematic function h^* of observable exogenous variables nor the distribution F^* of the random term is parametric.

For each of these models, I show that although h^* is not identified, it is possible to estimate a least concave representative of h^*, which possesses the same isovalue sets as h^*. The computation technique presented in the previous section can be applied to calculate the estimator of the least concave representative of h^* also in these models.

A generalized regression model is studied next. Section 7 deals with the binary threshold crossing model.

6.1. THE MODEL

Generalized regression models include, as special cases, the linear regression model, Box and Cox transformations, proportional and additive hazard models, censored regression models, and threshold crossing models. (See Han (1987) for details.)

In generalized regression models, an observable real variable y obeys the relationship

$$y = G(\ h^*(x),\ \eta)\ ,$$

where $x \in X$ is a vector of observable exogenous variables and η is an unobservable random variable. The values of x and η are assumed to be distributed independently of each other with probability measures P_x and P_η, respectively. The function G is nonconstant and monotone increasing in each of its arguments.

I consider the particular case in which $h^*:X \to R$ belongs to the set $\tilde{W}$, of monotone and concave functions that possess least concave representatives in the set W defined in Section 3, and the functions P_η and G are such $h^*(x_1) < h^*(x_2)$ implies that for some t, $P_{\eta|x_1}[G(h^*(x_1),\eta) \leq t] > P_{\eta|x_2}[G(h^*(x_2),\eta) \leq t]$.[4] The functions h^* and G and the distribution functions of x and η are assumed to be otherwise unknown.

In this model it is not possible to identify h^* from observations on the vector (y,x). To see this, note that the most we can observe from this data is the probability of observing any value of y smaller than t given x. A function h is observationally equivalent to h^* if there exist functions G' and P'_η possessing the same properties that G and P_η are known to possess and are such that, for any t, the probability of observing any value of y smaller than t given G', P'_η, and h is the true probability of this event. Formally,

DEFINITION 5: *The function* $h \in \tilde{W}$ *is* *<u>observationally equivalent</u>* *to* h^* *if there exists a monotone increasing, nonconstant function* G' *and a distribution* P'_η *such that*

(i) $\forall\ x_1, x_2 \in X$ $[\ h(x_1) < h(x_2)$ *implies that for some* t

$$P'_{\eta|x_1}[\ G(h(x_1),\eta) \le t\ ] > P'_{\eta|x_2}[\ G(h(x_2),\eta) \le t\ ] \quad \text{and}$$

(ii) $\forall\ x \in X$ *and all values* t *of* y

$$P_{\eta|x_1}[\ G(h^*(x_1),\eta) \le t\ ] = P'_{\eta|x_2}[\ G'(h(x_2),\eta) \le t\ ].$$

As the next lemma shows, there exist functions in $\tilde{W}$ that are observationally equivalent to h^*.

LEMMA 7: *There exists a concave and monotone function* $h \in \tilde{W}$ *that is observationally equivalent to* h^* *and is such that* $h \ne h^*$.

Nevertheless, it is possible to characterize each class of observationally equivalent functions in $\tilde{W}$ by a unique function in W.

THEOREM 6: *Suppose that* $h, h^* \in \tilde{W}$. *Let* $\bar{h}$ *and* $\bar{h}^*$ *denote respectively the least concave representatives in* W *of* h *and* h^*. *Then,* h *is observationally equivalent to* h^* *if and only if* $\bar{h} = \bar{h}^*$.

6.2. ESTIMATION

As with the binary choice model previously studied, in this model we can also develop an estimator for $\bar{h}^*$. To obtain such an estimator, we modify the Maximum Rank Correlation Method (see Han (1987) and Matzkin (1990a)). Han developed this method to estimate the function h^* in a generalized regression model where h^* was linear in a finite dimensional parameter β^*. Matzkin (1990a) modified Han's method to estimate a nonparametric function h^*.

Following Han (1987) and Matzkin (1990a), we define the rank correlation function by

(20) $S_N(z^{(N)},h) =$

$$= \binom{N}{2}^{-1} \Sigma_\rho \left\{ 1[h(x^i) > h(x^j)]\ 1[y^i > y^j] + 1[h(x^i) < h(x^j)]\ 1[y^i < y^j] \right\}$$

where $1[\cdot]$ is a logical operator that equals 1 if $[\cdot]$ is true and 0 otherwise, and Σ_ρ denotes the summation over the $\binom{N}{2}$ combinations of two distinct elements (i,j) from $\{1,\ldots,N\}$.

We define the *least concave maximum rank correlation* estimator to be any function $\hat{h}_N$: $X \to R$ that maximizes $S_N(z^{(N)},\cdot)$ over the set W.

Theorem 7 and 8 below show that this estimator is strongly consistent when the following assumptions are satisfied:

ASSUMPTION B.1: For all i,j, η^i is independent of η^j.

ASSUMPTION B.2: For all i η^i is independent of x^i.

ASSUMPTION B.3: For all i,j, x^i is independent of x^j.

ASSUMPTION B.4: P_x is absolutely continuous and its support is X.

ASSUMPTION B.5: For all $x^i, x^j \in X$ such that $h^*(x^i) < h^*(x^j)$, there exists $t^* \in R$ such that $P_{\eta|x}[\, y^i \le t^* \,] > P_{\eta|x}[\, y^i \le t^* \,]$, where $P_{\eta|x}$ denotes the probability with respect to η^i conditional on x^i.

ASSUMPTION B.6: $G{:}h^*(X) \to R$ is monotone increasing in each coordinate and nonconstant.

ASSUMPTION B.7: $h^* \in \tilde{W}$.

THEOREM 7 (Identification): *Suppose that Assumptions B.1-B.7 are satisfied. Then,* $\bar{h}^*$ *uniquely maximizes the expectation of the rank correlation function* $S_N(z^{(N)}, h)$.

THEOREM 8 (Consistency): *Suppose that Assumptions B.1-B.7 are satisfied. Then,* $\lim_{N\to\infty} d(\hat{h}_N, \bar{h}^*) = 0$ *a.s.*

Since the rank correlation function (20) is order-dependent, the computation of $\hat{h}_N$ can be performed using the technique presented in Section 5.

7. A THRESHOLD CROSSING MODEL

7.1. THE MODEL

Threshold crossing models have been applied to a variety of problems in economics, medicine, and other fields. Subjects studied include labor force participation, acceptance of loan applications, and health status.

In this model an observable dichotomous variable y is determined by

$$y = 1 [h^*(x) - \eta \geq 0] ,$$

where $1[\cdot]$ is the logical operator that equals 1 if $[\cdot]$ is true and 0 otherwise, x is a vector of observable exogenous variables in X, and η is an unobservable random term. The vector x possesses a probability measure P_x whose support is the closure of X.

In the loan application acceptance model, for example, x may denote assets and income of the applicant and h^* the expected value to the bank of accepting his loan application; y equals one if the application is accepted and y equals zero otherwise.

I consider the particular case of this model in which the function h^* belongs to $\tilde{W}$, the set of all concave functions that posses least concave representatives in W , and the random term η is independent of x and possesses a strictly increasing and continuous distribution function F^* . The functions h^* and F^* are otherwise unknown.

As I show below, also in this model h^* can not be identified. Denote the probability that $y = j$ given x by $P(j| x; h^*,F^*)$. And note that $P(j| x; h^*,F^*)$ $(j=1,2)$ is the most we can obtain from the data. The definition of the model implies that $P(1| x; h^*,F^*) = F^*(h^*(x))$ and $P(2| x; h^*,F^*) = 1 - P(1| x; h^*,F^*)$. Let $\Gamma = \{ F: [\alpha,\gamma] \to [0,1] \mid F \text{ is monotone increasing} \}$. Then,

DEFINITION 6: *The pair $(h,F) \in (\tilde{W}\times\Gamma)$ is <u>observationally equivalent</u> to the pair (h^*,F^*) if for all $x \in X$ $P(j| x; h^*,F^*) = P(j| x; h,F)$ $j=1,2$.*

LEMMA 8: *There exists a pair* $(h,F) \in (\tilde{W} \times \Gamma)$ *that is observationally equivalent to* (h^*,F^*) *and is such that* $(h,F) \neq (h^*,F^*)$.

Hence, in this model we can not identify (h^*,F^*) within $(\tilde{W} \times \Gamma)$. Nevertheless, the next theorem shows that we can characterize the set of pairs (h,F) that are observationally equivalent to (h^*,F^*) in $(\tilde{W} \times \Gamma)$.

LEMMA 9: *Suppose that* $(h,F) \in (\tilde{W} \times \Gamma)$. *Let* $\bar{h}$ *and* $\bar{h}^*$ *denote respectively the least concave representatives of* h *and* h^*. *Define* $\bar{F}$ *and* $\bar{F}^*$, *respectively, by: for all* $x \in X$, $\bar{F}(\bar{h}(x)) = F(h(x))$ *and* $\bar{F}^*(\bar{h}^*(x)) = F^*(h^*(x))$. *Then,* (h,F) *is observationally equivalent to* (h^*,F^*) *if and only if* $\bar{h} = \bar{h}^*$ *and* $\bar{F} = \bar{F}^*$.

From an estimator for $(\bar{h}^*,\bar{F}^*)$, we can obtain estimator of the equivalence class of (h^*,F^*).

$$\{ (h,F) \in (\tilde{W} \times \Gamma) \mid (h,F) \text{ is observationally equivalent to } (h^*,F^*) \}$$
$$= \{ (h,F) \mid \text{for some concave and strictly increasing function } g: [\alpha,\gamma] \to R,\ h = g \circ \bar{h}^* \text{ and } F = \bar{F}^* \circ g^{-1} \}.$$

In the next subsection I show how an estimator for $(\bar{h}^*, \bar{F}^*)$ can be obtained.

7.2. ESTIMATION

To develop an estimator for $(\bar{h}^*,\bar{F}^*)$, I follow Cosslett (1983) and Matzkin (1988). Define the conditional log-likelihood function of a sample of N independent observations $z^{(N)} = \{y^i, x^i\}_{i=1}^N$ at any function $h: X \to R$ and monotone increasing function $F: R \to [0,1]$ by

$$(21)\quad \mathcal{L}(z^{(N)}, h, F) = \sum_{i=1}^{N} \{ y^i \log(F(h(x^i)) + (1 - y^i) \log(1 - F(h(x^i)) \}$$

And for any function $h: X \to R$, define the *concentrated log-likelihood function* by

$$(22)\quad S_N (z^{(N)}, h) = \max_{F\in\Gamma} \mathcal{L} (z^{(N)}, h, F) .$$

The *least concave maximum likelihood* estimator is defined to be the pair $(\hat{h}_N, \hat{F}_N)$ such that $\hat{h}_N: X \to R$ maximizes $S_N(z^{(N)}, \cdot)$ over the set W and $\hat{F}_N: R \to [0,1]$ maximizes $\mathcal{L}(z^{(N)}, \hat{h}_N, \cdot)$ over Γ.

Theorems 10 and 11 show that this estimator is strongly consistent when the following assumptions are satisfied:

ASSUMPTION C.1: The random term η is distributed independently of x with a cumulative distribution function F^*.

ASSUMPTION C.2: F^* is strictly increasing.

ASSUMPTION C.3: The vector x possesses a Lebesgue density g whose support is X and whose probability measure is P_x.

ASSUMPTION C.4: The probability density g is bounded.

ASSUMPTION C.5: $h^* \in \tilde{W}$.

THEOREM 10 (Identification): *Suppose that Assumptions C.1-C.5 are satisfied. If $(h,F) \in (W\times\Gamma)$ and $(h,F) \neq (\bar{h}^*, \bar{F}^*)$, then for some $A \subset X$,*

$$\int_A g(x)\ F(h(x))\ dP_x(x) \neq \int_A g(x)\ \bar{F}^*(\bar{h}^*(x))\ dP_x(x) .$$

THEOREM 11 (Consistency): *Suppose that Assumptions C.1-C.5 are satisfied. Then, with probability one,*

$$lim_{N\to\infty}\ d(\hat{h}_N,\ \bar{h}^*) = 0 \quad \text{and} \quad lim_{N\to\infty}\ d_\Gamma(\hat{F}_N,\ \bar{F}^*) = 0\ ,$$

where the metric $d_\Gamma\colon \Gamma \times \Gamma \to R$ *is defined by*

$$d_\Gamma(F,\ F') = \int \ |F(t) - F'(t)| \quad dt\ ,$$

(the integration is with respect to the Lebesgue measure over $[\alpha,\gamma]$*)*[5].

Hence a strongly consistent estimator for the unique representative of (h^*, F^*) can be obtained by maximizing (21) over $(W \times \Gamma)$.

The function $S_N(z^{(N)},h)$ in (22) is order-dependent (see Cosslett (1983)). Hence, the maximization of (22) over W can be performed by following the technique introduced in Section 5.

8. CONCLUSION

I have considered the problem of estimating fully nonparametric models in which the values of the systematic is not identified. In these models, neither the systematic function of observable exogenous variables nor the distribution of the random term possesses a parametric structure. I have shown that, when the systematic function is monotone and concave, it is possible to obtain a consistent nonparametric estimator for a representative of this systematic function. The isovalue sets of this representative coincide with the isovalue sets of the systematic function. The method has been described in three microeconometric models.

The estimation method proceeds by characterizing each set of observationally equivalent functions with a unique least concave representative in a compact set of least concave functions. Given its least concave representative, the entire equivalence class can be obtained by composing this representative with monotone increasing and concave functions.

I have introduced a technique of estimating the values, at each observed point, and the gradients, at a.e. point and up to a positive constant, of the least concave representative of the systematic function. This technique proceeds in two steps. In the first step, a criterion function is maximized over a set of monotone increasing and concave functions. In the second step, the values, at the observed vectors of exogenous variables, and the gradients, at a.e. point, of a least concave function are obtained.

The three models that have been studied are a binary choice model, a generalized regression model, and a binary threshold crossing model. The random term in the first model is median independent of the exogenous observable regressors, while the random terms in the second and third models are independent of the regressors. In the first and second models, I have concentrated on estimating the representative of the systematic function. In the third model, I have studied the estimation of the representatives of both the systematic function and the distribution of the random term. In all these models, I have shown that the proposed estimators for the least concave representatives are strongly consistent.

Summing up, the method introduced in this paper provides a way of estimating fully nonparametric models in which the restrictions on the functions and distributions are not strong enough to be able to identify the systematic function generating the observations.

APPENDIX

PROOF OF LEMMA 1: Let f be a strictly concave and strictly increasing function, and define h by $h = f \cdot h^*$. Then, h is concave, monotone increasing, and $h = f \cdot h^* \neq h^*$.

We next show that since f is strictly increasing, $f \cdot h^*$ and h^* are observationally equivalent. Take any function F of (η, x) such that for each $x = (x_1, x_2)$, (i) $F(\cdot, x)$ is a monotone increasing function of η, (ii) $F(0,x) = .5$, and (iii) the value of $F(\cdot, x)$ at $[f \cdot h^*(x_1) - f \cdot h^*(x_2)]$ coincides with the value of $F^*_{\eta|x}$ at $[h^*(x_1) - h^*(x_2)]$, i.e.

(L1.1) $F(\ f \cdot h^*(x_1) - f \cdot h^*(x_2),\ x_1, x_2\ ) = F^*_{\eta|x}\ (h^*(x_1) - h^*(x_2))$.

(To see that such a function F always exists, note that since f is strictly increasing,

$$[\ f \cdot h^*(x_1) - f \cdot h^*(x_2) \geq 0\ ] \iff [\ h^*(x_1) - h^*(x_2) \geq 0\ ].$$

This, (L1.1), and the fact the $F^*_{\eta|x}(0) = .5$ imply that

$[\ F(\ f \cdot h^*(x_1) - f \cdot h^*(x_2),\ x_1, x_2) \geq .5\ ]$ if $[\ f \cdot h^*(x_1) - f \cdot h^*(x_2) > 0]$ and

$[\ F(\ f \cdot h^*(x_1) - f \cdot h^*(x_2),\ x_1, x_2) \leq .5\ ]$ if $[\ f \cdot h^*(x_1) - f \cdot h^*(x_2) < 0]$,

which will always be consistent with the requirement that $F(\cdot, x)$ be monotone increasing and $F(0,x) = .5$)

Define $F_{\eta|x}$ by $F_{\eta|x}(\eta) = F(\eta', x)$ for all $x \in (X \times X)$. Then, for all $x \in (X \times X)$

$$\begin{aligned} P(1|\ x_1, x_2;\ h, F) &= F(\ f \cdot h^*(x_1) - f \cdot h^*(x_2),\ x_1, x_2) \\ &= F^*_{\eta|x}(\ h^*(x_1) - h^*(x_2)) \\ &= P(1|\ x_1, x_2;\ h^*, F^*). \end{aligned}$$

Hence, h is observationally equivalent to h^*. Q.E.D.

PROOF OF LEMMA 2: Suppose that $h' = f \cdot h$ for some strictly increasing function f. To see that $\bar{h}$ is a least concave representative of h', let s be strictly increasing. Then, $s \cdot h' = s \cdot f \cdot h$ is a strictly increasing transformation of h. It follows that for some concave function g, $s \cdot h' = g \cdot \bar{h}$. Hence, $\bar{h}$ is a least concave representative of h'.

Next, if $\bar{h}$ is a least concave representative of h and h', then there exist strictly increasing (and concave) functions g and s such that $h = g \cdot \bar{h}$ and $h' = s \cdot \bar{h}$. Thus, the strictly increasing function $f = s \cdot g^{-1}$ satisfies $h' = f \cdot h$.

PROOF OF LEMMA 3: Define the binary relation $\gtrsim$ on X by: $\forall x,y \in X$ $x \gtrsim y \iff h(x) \geq h(y)$. Then, $\gtrsim$ possesses a concave representation. It follows by Debreu (1976, Theorem) that there exists $\bar{h}$ such that $\bar{h}$ is a least concave representative of h. Q.E.D.

PROOF OF LEMMA 4: See Debreu (1976).

PROOF OF LEMMA 5: Since $\bar{h}$ is a least-concave representative of h, there exists a concave function g such that $h = g \cdot \bar{h}$. We will show that g is linear. Suppose otherwise that there exists $\bar{h}_1$, $\bar{h}_2$, $\bar{h}_3$ in $\bar{h}(X)$ with $\bar{h}_1 < \bar{h}_2 < \bar{h}_3$ and such that

$$g(\bar{h}_2) > \frac{\bar{h}_3 - \bar{h}_2}{\bar{h}_3 - \bar{h}_1} g(\bar{h}_1) + \frac{\bar{h}_2 - \bar{h}_1}{\bar{h}_3 - \bar{h}_1} g(\bar{h}_3) ,$$

or equivalently,

$$\frac{g(\bar{h}_2) - g(\bar{h}_1)}{g(\bar{h}_3) - g(\bar{h}_1)} > \frac{\bar{h}_2 - \bar{h}_1}{\bar{h}_3 - \bar{h}_1} .$$

Let x^1, x^2 and x^3 be three collinear elements in X such that $h(x^i) = g(\bar{h}_i)$ $(i = 1, 2, 3)$. (The existence of these elements is guaranteed by the continuity of h and the convexity of X.) Since h is linear and $\bar{h}$ is concave it follows that

$$\frac{\|x^2 - x^1\|}{\|x^3 - x^1\|} = \frac{h(x_2) - h(x_1)}{h(x_3) - h(x_1)} = \frac{g(\bar{h}_2) - g(\bar{h}_1)}{g(\bar{h}_3) - g(\bar{h}_1)} > \frac{\bar{h}_2 - \bar{h}_1}{\bar{h}_3 - \bar{h}_1} \geq \frac{\|x^2 - x^1\|}{\|x^3 - x^1\|}$$

which is a contradiction. Hence, g is linear. Then, since $\bar{h} = (g)^{-1} \cdot h$, $\bar{h}$ is linear. Q.E.D.

PROOF OF LEMMA 6: See Debreu (1976, Proposition) or Benveniste and Scheinkman (1979).

PROOF OF THEOREM 1: The set of C^2 functions with uniformly Lipschitzian and uniformly bounded values, gradients, and Hessians is compact with respect to d. Hence, it is only necessary to show that any convergent sequence of monotone, concave, and least concave functions in W that satisfy (ii) converge to a function satisfying these same properties. It is clear that the limiting function is concave, monotone, and satisfies (ii). Moreover, the limiting function possesses Gaussian curvatures uniformly bounded below by c, because its gradients and Hessians are limits of the gradients and Hessians of the functions in the sequence. By Kannai (1977)[6], the limit of a convergent sequence of least concave functions in W converges to a least concave function. Hence, the limiting function satisfies the desired properties.

Q.E.D.

PROOF OF THEOREM 2: Suppose that h is observationally equivalent to h^*. We show that then there exists a strictly increasing function f such that $h = f \cdot h^*$. Suppose that such a function does not exist. Then, we must be able to find some $(x_1, x_2) \in (X \times X)$ such that either

$[\ h(x_1) > h(x_2)$ and $h^*(x_1) \le h^*(x_2)\ ]$ or

$[\ h(x_1) \le h(x_2)$ and $h^*(x_1) > h^*(x_2)\ ]$.

Suppose w.l.o.g. that the first case is true, Then, since both h and h^* are continuous and strictly increasing with respect to the Kth coordinate, there exists $(x_1', x_2') \in (X \times X)$ such that $h(x_1') > h(x_2')$ and $h^*(x_1') < h^*(x_2')$. It then follows that for any strictly increasing distribution function $F_{\eta|x}$ such that $F_{\eta|x}(0) = .5$:

$$F_{\eta|x}(h(x_1') - h(x_2')) > .5 \quad \text{while} \quad F^*_{\eta|x}(h^*(x_1') - h^*(x_2')) < .5$$

Thus,

$$P(1|x_1', x_2'; h^*, F^*) < P(1|x_1', x_2'; h, F),$$

which violates the hypothesis that h is observationally equivalent to h^*.

Hence, $h = f \cdot h^*$ for some strictly increasing function f. It follows by Lemma 2 that h and h^* possess the same least concave representatives. Since the least concave representative in W of any function in $\tilde{W}$ is unique, $\bar{h}^* = \bar{h}$.

Conversely, suppose that $\bar{h}^* = \bar{h}$. Then, by Lemma 2 there exist a strictly increasing function f such that $h = f \cdot h^*$. As shown in the proof of Lemma 1, this implies that h is observationally equivalent to h^*. Q.E.D.

In the proof of Theorem 3 we will make use of the following auxiliary lemma. This lemma will be used also in the proofs of theorems 7 and 9.

LEMMA 0: *Suppose that h and h' belong to the set of least concave functions W defined in Section 3 and $h \neq h'$. Then, there exist x_1' and x_2' in X, and neighborhoods $X_1 \subset X$ and $X_2 \subset X$ of x_1' and x_2', respectively, such that for all $x_1 \in X_1$ and $x_2 \in X_2$,*

$$h(x_1) < h(x_2) \quad \text{and} \quad h'(x_1) > h'(x_2) .$$

PROOF OF LEMMA 0: We first show that there exists no strictly increasing function f such that $h = f \cdot h'$. Suppose such a function f exists. Then h is a least concave representative of h' and h' is a least concave representative of h. This is possible only if f is linear. But if f is linear, Lemma 4 and the definition of W imply that $h = h'$, a contradiction. Hence, there does not exist a strictly increasing function f such that $h = f \cdot h'$. Thus, there must exist x_1' and x_2' in X such that either

$$h(x_1') \geq h(x_2') \quad \text{and} \quad h'(x_1') < h'(x_2'), \quad \text{or}$$

$$h(x_1') > h(x_2') \quad \text{and} \quad h'(x_1') \leq h'(x_2').$$

By the definition of X and the continuity and monotonicity of h and h'

it follows that x_1' and x_2' can be assumed to belong to int(X). Then, the continuity of h and h' and the strict monotonicity of h and h' in the Kth coordinate imply that there exist neighborhoods, $X_1 \subset X$ and $X_2 \subset X$ of x_1' and x_2', respectively, such that for all $x_1 \in X_1$ and $x_2 \in X_2$,

$$h(x_1) > h(x_2) \quad \text{and} \quad h'(x_1) < h'(x_2).$$

Q.E.D.

PROOF OF THEOREM 3: For any $h \in W$ such that $h \neq \bar{h}^*$, let

$$X_h = \{ (x_1, x_2) \in (X \times X) \mid \text{sgn}(h(x_1) - h(x_2)) \neq \text{sgn}(\bar{h}^*(x_1) - \bar{h}^*(x_2)) \}.$$

Then, from the definition of $S(\cdot)$ it follows that

$$S(\bar{h}^*) - S(h) = E [y [\text{sgn}(\bar{h}^*(x_1) - \bar{h}^*(x_2)) - \text{sgn}(h(x_1) - h(x_2))]]$$

$$= 2 \int_{X_h} E[y \mid x_1, x_2] \ \text{sgn}(\bar{h}^*(x_1) - \bar{h}^*(x_2)) \ dP_x ,$$

Since

$$\text{MEDIAN}[y^* \mid x_1, x_2] = h^*(x_1) - h^*(x_2) ,$$
$$\text{sgn}(\bar{h}^*(x_1) - \bar{h}^*(x_2)) = \text{sgn}(h^*(x_1) - h^*(x_2)) , \text{ and}$$
$$E[y \mid x_1, x_2] = 2 \Pr(y^* \geq 0 \mid x) - 1 ,$$

it follows that

$$\text{sgn}(\bar{h}^*(x_1) - \bar{h}^*(x_2)) = \text{sgn } E[y \mid x] .$$

Thus, for all $x \in X$, $E[y \mid x] \ \text{sgn}(\bar{h}^*(x_1) - \bar{h}^*(x_2)) = | E[y \mid x] |$.
It then follows that $S(\bar{h}^*) - S(h) > 0$ if

$$\text{(T3.1)} \qquad \int_{X_h} dP_x > 0 .$$

To show that (T3.1) is satisfied, we note that since h and $\bar{h}^*$ belong to W and $h \neq \bar{h}^*$, Lemma 0 implies that there exist neighborhoods X_1 and X_2 in int(X) such that for all $x_1 \in X_1$ and $x_2 \in X_2$,

$$\bar{h}^*(x_1) - \bar{h}^*(x_2) > 0 \text{ and } h(x_1) - h(x_2) < 0.$$

Since by our assumptions on P_x, $P_x(X_1 \times X_2) > 0$, it follows that

$$\int_{X_h} dP_x \geq \int_{X_1 x X_2} dP_x > 0 .$$

Hence, for all $h \in W$ such that $h \neq \bar{h}^*$, $S(\bar{h}^*) > S(h)$. Q.E.D.

PROOF OF THEOREM 4: We will show the statement of the theorem in a sequence of steps.

Step 1 (Uniform Convergence): $sup_{h \in W} \mid S_N(h) - S(h) \mid \to 0$ *a.s.*

We will show that with probability one,

(T4.1) $\sup_{h\in W} \mid P_N[y \geq 0,\ h(x_1)-h(x_2) \geq 0] - P[y \geq 0,\ h(x_1)-h(x_2) \geq 0] \mid \to 0,$

where P_N is the empirical distribution of (y,x_1,x_2). The uniform convergence of the other terms of the sample score function $S_N(\cdot)$ can be shown similarly.

For any $h \in W$, h is monotone increasing and concave. Hence,

$\{ (x_1,x_2) \in X \times X \mid h(x_1) \geq h(x_2) \} =$

$\{ (x_1, x_2) \mid x_1 \in C_1,\ x_2 \in C_2$ for some monotone increasing[7] and convex sets C_1, C_2 such that $C_1 \subset C_2 \}$.

Thus,

(T4.2) $\{(x_1,x_2) \in X \times X \mid h(x_1) \geq h(x_2) \} \subset P^K \times P^K$

where P^K is the set of all convex subset of X. Since P^K does not depend on h, (T4.2) holds for all h in W. Moreover, since $P^K \times P^K \subset P^{2K}$, the set of all convex sets in R^{2K},

$$\sup_{h\in W} \left| P_N[y^* \geq 0,\ h(x_1) - h(x_2) \geq 0] - P[y^* \geq 0,\ h(x_1) - h(x_2) \geq 0] \right|$$

$$\leq \sup_{C_1\times C_2 \in P^K \times P^K} \left| P_N[y^* \geq 0,\ (x_1,x_2) \in C_1\times C_2\ ] - P[y^* \geq 0,\ (x_1,x_2) \in C_1\times C_2\ ] \right|$$

$$\leq \sup_{C \in P^{2K}} \left| P_N[y^* \geq 0,\ (x_1,x_2) \in C\times C)\ ] - P(y^* \geq 0,\ (x_1,x_2) \in C\times C)\ ] \right|.$$

By Rao (1962, Theorem 7.1), the last term converges to 0 a.s. Hence, (T4.1) follows. Q.E.D.

Step 2 (Continuity of $S(\cdot)$): *$S(\cdot)$ is a continuous function on W.*

Denote x_1 by $x_1 = (\tilde{x}_1, x_K)$, where x_K denotes the Kth coordinate of x_1. Let $P_{(\tilde{x}_1,x_2)}$ denote the probability measure of $(\tilde{x}_1,x_2) \in R^{(K-1)\times K}$ and let $f(x_K|\tilde{x}_1,x_2)$ denote the probability density of x_K conditional on $(\tilde{x}_1,x_2)$. By Assumption A.3, $f(x_K|\tilde{x}_1,x_2)$ is a Lebesgue density.

We will show that the term $P[y^* \geq 0, h(x_1)-h(x_2) \geq 0]$ is continuous in h. The continuity of the other terms in $S(\cdot)$ can be shown in a similar way.

Let $(\tilde{x}_1,x_2)$ be given. Suppose that $h_k \to h$ with respect to d. Define functions r_k $(k=1,2,\ldots)$ and r by $r_k(x_K) = h_k(\tilde{x}_1,x_K)$ and $r(x_K) = h(\tilde{x}_1,x_K)$ for all x_K such that $(\tilde{x}_1,x_K)\in X$. These functions are strictly increasing and continuous and $\sup | r_k(x_K) - r(x_K) | \to 0$. Our aim is to show that $r_k^{-1}(h_k(x_2)) \to r^{-1}(h(x_2))$.

Let $\varepsilon > 0$ be given. Since $r(\cdot)$ is strictly increasing and continuous and X is compact, there exists $\delta > 0$ such that for all x_K, x_K' such that $(\tilde{x}_1, x_K), (\tilde{x}_1, x_K') \in X$

(T4.3) $\quad | x_K - x_K' | \geq \varepsilon \implies | r(x_K) - r(x_K') | > \delta.$

To see this, note that if $x_K - x_K' \geq \varepsilon$, $r(x_K) - r(x_K') \geq r(x_K'+\varepsilon) - r(x_K') > 0$. Hence, since $r(x_K'+\varepsilon) - r(x_K')$ is a continuous function on a compact set, it must attain a minimum, strictly positive, value on its domain. Similarly, if $x_K - x_K' \leq \varepsilon$, $r(x_K) - r(x_K') \leq r(x_K'+\varepsilon) - r(x_K') < 0$. So, $r(x_K'+\varepsilon) - r(x_K')$ attains a maximum, strictly negative, value on its domain.

Since $\sup | r_k(x_K) - r(x_K) | \to 0$, $| r_k(x_K) - r(x_K) | \leq \delta/4$ for all x_K and all large enough k. Hence,

$$\text{if } | r_k(x_K) - r(x_K') | \leq \delta/4, \quad | r(x_K) - r(x_K') | \leq \delta.$$

By (T4.3) it then follows that $| x_K - x_K' | < \varepsilon$. So,

$$| r_k(x_K) - r(x_K') | \leq \delta/4 \implies | x_K - x_K' | < \varepsilon.$$

In particular, if x_K and x_K' are such that $r_k(x_K) = h_k(x_2)$ and $r(x_K') = h(x_2)$ and k is large enough such that $| h_k(x_2) - h(x_2) | \leq \delta/4$,

$$| r_k^{-1}(h_k(x_2)) - r^{-1}(h(x_2) | < \varepsilon.$$

Let $r^{-1}(t)=\bar{x}_K$ if for all $x_K \in [\underline{x}_K, \bar{x}]$ $r(x_K) \leq t$ and $r^{-1}(t)=\underline{x}_K$ if for all $x_K \in [\underline{x}_K, \bar{x}]$ $r(x_K) \geq t$; let $r_k^{-1}(t)$ be defined similarly. Then, it follows that if $h_k \to h$ with respect to d,

$$r_k^{-1}(h_k(x_2)) \to r^{-1}(h(x_2) \quad \text{for all } x_2 \in X.$$

We next note that

$$P[\, y^* \geq 0,\ h(x_1) - h(x_2) \geq 0 \,]$$

$$= \int_{(\tilde{x}_1, x_2)} \left[\int_{r^{-1}(h(x_2))}^{\bar{x}_K} P[\, y^* \geq 0 \mid x \,] \; f(x_K \mid \tilde{x}_1, x_2) \; dx_K \right] dP_{(\tilde{x}_1, x_2)} .$$

Since the term in brackets is measurable, uniformly bounded, and, by the above argument, continuous in h, it follows from the Lebesgue Dominated Convergence Theorem that $P[\ y \geq 0,\ h(x_1) - h(x_2) \geq 0\ ]$ is continuous in h.

Q.E.D.

Step 3 (Consistency):

This follows from the compactness of W (Theorem 1), Theorem 3, and Steps 2 and 3. (See, for example, the proof of Step 4 in Matzkin (1990a).)

Q.E.D.

PROOF OF THEOREM 5: Define a function $h: X \to R$ by

(T5.1) $h(x) = \min \{ h_i^* + \beta_i^* (x - x^i) \mid i=0,1,\ldots,N+1 \}.$

Then, h is a monotone increasing, concave, strictly increasing in the Kth coordinate, and such that

(T5.2) $h(x^i) = h_i^*$ $(i= 0,\ldots,N+1).$

(See Matzkin (1987).) By Lemma 3, there exists a least concave function $\bar{h}$ that represents h. By Lemma 4 we can assume that $\bar{h}(\underline{x}) = \alpha$ and $\bar{h}(\bar{x}) = \gamma$. The function $\bar{h}$ is monotone increasing, concave, and strictly increasing in the Kth coordinate. Moreover, since by (T5.2), (6), and (7), h induces the same preorder on $\{x^0,\ldots,x^{N+1}\}$ as g does, it follows that $\bar{h}$ also induces this same preorder on $\{x^0,\ldots,x^{N+1}\}$. This proves (i).

Next, we show that

(T5.3) $\bar{h}(x^i) = h_i^*$ $(i=1,\ldots,N).$

Suppose first that for some $i \in \{1,\ldots,N+1\}$ $\bar{h}(x^i) > h_i^*$. Since $\bar{h}$ is a least concave representative of the function h defined in (T5.1), $h = f \circ \bar{h}$ for some concave function $f: \bar{h}(X) \to R$. By (T5.2) and the fact that $\bar{h}(\underline{x}) = \alpha$ and $\bar{h}(\bar{x}) = \gamma$, $f(\alpha) = \alpha$ and $f(\gamma) = \gamma$. Since f is concave, $\bar{h}(x^i) \in [\alpha, \gamma]$, and α and γ belong to $\bar{h}(X)$,

$$f(\bar{h}(x^i)) \geq \frac{\bar{h}(x^i) - \alpha}{\gamma - \alpha} f(\bar{h}(\bar{x})) + \frac{\gamma - \bar{h}(x^i)}{\gamma - \alpha} f(\bar{h}(\underline{x})).$$

Hence,

$$\frac{f(\bar{h}(x^i)) - \alpha}{\gamma - \alpha} \geq \frac{\bar{h}(x^i) - \alpha}{\gamma - \alpha}.$$

But then, since $\bar{h}(x^i) > h_i^* = h(x^i)$ and $f(\bar{h}(x^i)) = h(x^i)$,

$$\frac{h(x^i) - \alpha}{\gamma - \alpha} = \frac{f(\bar{h}(x^i)) - \alpha}{\gamma - \alpha} \geq \frac{\bar{h}(x^i) - \alpha}{\gamma - \alpha} > \frac{h(x^i)) - \alpha}{\gamma - \alpha},$$

which is a contradiction. Hence,

(T5.4) $\bar{h}(x^i) \leq h_i^*$, for all $i=1,\ldots,N+1$.

Suppose now that for some i, $\bar{h}(x^i) < h_i^*$. By (T5.4) this implies that

(T5.5) $$\Sigma_{j=1}^N p_j \bar{h}(x^j) < \Sigma_{j=1}^N p_j h(x^j) = \Sigma_{j=1}^N p_j \bar{h}_j^*.$$

But, the properties of $\bar{h}$ imply that $\bar{h}(x^1),\ldots,\bar{h}(x^N)$ satisfy (4)-(8) for some $\beta_0,\ldots,\beta_{N+1}$. Hence, (T5.5) contradicts the fact that $h_1^*,\ldots,h_N^*$ solves the minimization problem in (3)-(8). This contradiction implies that

(T5.6) $\bar{h}(x^i) \geq h_i^*$, for all $i=1,\ldots,N$.

From (T5.4) and (T5.6), $\bar{h}(x^i) = h_i^*$ for all $i = 1,\ldots,2N+1$. This proves (T5.3) and hence, (ii) is proved.

To prove (iii), we note that since $h = f \circ \bar{h}$ for some strictly increasing and concave function f, it follows that at any $x \in X$ such that $\bar{h}$ is differentiable at x and f is differentiable at $\bar{h}(x)$, one has that $Dh(x) = Df(\bar{h}(x)) \cdot D\bar{h}(x)$, where $Df(\bar{h}(x)) > 0$ since f is strictly increasing. Hence, at any such x

(T5.7) $D\bar{h}(x) = a(x)\ Dh(x)$

for $a(x) = (1/Df(\bar{h}(x))) > 0$.

We next show that the set of x's at which $f \circ \bar{h}$ is not differentiable has Lebesgue measure zero. To prove this, we note that concave functions are differentiable a.e. Hence, the set, $d(f)$, of values t at which f is not differentiable has Lebesgue measure zero. Since $\bar{h}$ is continuous and strictly increasing on its Kth coordinate, the set of all points x at which $\bar{h}(x) \in d(f)$ has Lebesgue measure zero. Hence, the set of points x at which f is not differentiable at $\bar{h}(x)$ has Lebesgue measure zero. From this latter result it follows that, since the set of points at which $\bar{h}$ is not differentiable has Lebesgue measure zero, the set of points at which (T5.7) is not satisfied has Lebesgue measure zero.

The result of (iii) now follows by (T5.1).

Q.E.D.

PROOF OF LEMMA 7: Let f be a C^2 strictly increasing function such that $f \circ h^*$ is concave and $f \circ h^* \neq h^*$ (take, for example, a strictly concave f). Let $h = f \circ h^*$. Then, $h \in \hat{W}$. Define the function $r:(h(X) \times R) \to R^2$ by $\forall (t,\eta) \in (h(X) \times R)$, $r(t,\eta) = (f^{-1}(t),\eta)$, and let $G' = G \circ r$. Then,

(L7.1) G' is nonconstant and monotone increasing in each argument and

(L7.2) $\forall (x,\eta)$ $G(h^*(x),\eta) = G'(h(x),\eta)$.

Let x_1, x_2 be such that $h(x_1) < h(x_2)$. Then, since h is a monotone increasing transformation of h^*, $h^*(x_1) < h^*(x_2)$. It then follows from our assumptions on G and P_η that for some t,

$$P_{\eta|x_1}(G(h^*(x),\eta) \le t) > P_{\eta|x_2}(G(h^*(x),\eta) \le t).$$

By (L7.2) this implies that

$$P_{\eta|x_1}(G'(h(x),\eta) \le t) = P_{\eta|x_1}(G(h^*(x),\eta) \le t)$$
$$> P_{\eta|x_2}(G(h^*(x),\eta) \le t) = P_{\eta|x_2}(G(h^*(x),\eta) \le t).$$

Hence,

(L7.3) $$P_{\eta|x_1}(G'(h(x),\eta) \le t) > P_{\eta|x_2}(G(h^*(x),\eta) \le t).$$

From (L7.2) it also follows that, for all y,

(L7.4) $$P_{\eta|x}(G(h^*(x),\eta) \le y) = P_{\eta|x}(G'(h(x),\eta) \le y).$$

Hence, by (L7.1), (L7.3), and (L7.4) it follows that h is observationally equivalent to h^*. Q.E.D.

PROOF OF THEOREM 6: Suppose that $h \in \hat{W}$ is observationally equivalent to h^*. We show that then there exists a strictly increasing function f such that $h = f \cdot h^*$. Suppose that such a function f does not exist. Then, there must exist $(x_1, x_2) \in (X \times X)$ such that either

$[\ h(x_1) > h(x_2)$ and $h^*(x_1) \le h^*(x_2)\ ]$ or

$[\ h(x_1) \le h(x_2)$ and $h^*(x_1) > h^*(x_2)\ ]$.

Suppose w.l.o.g. that $h(x_1) > h(x_2)$ and $h^*(x_1) \le h^*(x_2)$. Then, since both h and h^* are continuous and strictly increasing with respect to the Kth coordinate, there exists $(x_1', x_2') \in (X \times X)$ such that $h(x_1') > h(x_2')$ and $h^*(x_1') < h^*(x_2')$. Our assumptions on P_η and G imply then that for any monotone increasing and nonconstant function G' and any distribution P'_η, there exists some value t of y such that,

$$P'_{\eta|x'_1}(\ G'(h(x),\eta) \le t) \ \le \ P'_{\eta|x'_2}(\ G'(h(x),\eta) \le y) \quad \text{while}$$
$$P_{\eta|x'_1}(\ G(h^*(x),\eta) \le t) \ > \ P_{\eta|x'_2}(\ G(h^*(x),\eta) \le y).$$

This can only be possible if either

$$P'_{\eta|x'_1}(\ G'(h(x),\eta) \le t) \ne P_{\eta|x'_1}(\ G(h^*(x),\eta) \le t) \quad \text{or}$$
$$P'_{\eta|x'_2}(\ G'(h(x),\eta) \le t) \ne P_{\eta|x'_1}(\ G(h^*(x),\eta) \le t).$$

But, either of these possibilities contradicts the hypotheses that h is observationally equivalent to h^*. Hence, $h = f \cdot h^*$ for some strictly increasing function f. It follows by Lemma 2 that h and h^* possess common least concave representatives. Hence, Lemma 4 and condition (ii) in the definition of W imply that $\bar{h} = \bar{h}^*$.

Conversely, suppose that $\bar{h} = \bar{h}^*$. Then, by Lemma 2 there exists a strictly increasing function f such that $h = f \cdot h^*$. As can be seen from the proof of Lemma 7, this implies that h is observationally equivalent to h^*. Q.E.D.

PROOF OF THEOREM 7: For all $h \in W$, let

$$r_{ij}(h) = [1[h(x^i) > h(x^j)]\ 1[y^i > y^j] + 1[h(x^i) < h(x^j)]\ 1[y^i < y^j]] \ .$$

Then,

$$S_N(h) = \binom{N}{2}^{-1} \sum_\rho r_{ij}(h),$$

(T7.1) $E[S_N(h)] = E_x[\, E_{\eta|x} \binom{N}{2}^{-1} \Sigma_{\rho}\, r_{ij}(h) \,]$, and

$E_{\eta|x}[r_{ij}(h)] =$

$= 1[h(x^i) > h(x^j)]\; P_{\eta|x}(y^i > y^j) + 1[h(x^i) < h(x^j)]\; P_{\eta|x}(y^i < y^j)$

As is shown in Matzkin (1990a, Step 2, (2.a)), Assumptions B.1, B.2, B.5, and B.6 guarantee that

(T7.2) $\bar{h}^*(x^i) < \bar{h}^*(x^j)$ implies that $P_{\eta|x}(y^i > y^j) \leq P_{\eta|x}(y^i < y^j)$.

By Lemma 0 it follows that if $h \in W$ and $h \neq \bar{h}^*$ there exist neighborhoods X_i and X_j in X such that for all $x^i \in X_i$ and $x^j \in X_j$

(T7.3) $\bar{h}^*(x^i) < \bar{h}^*(x^j)$ and $h(x^i) > h(x^j)$;

and by Assumption B.3 the probability measure of $X_i \times X_j$ is strictly positive. Since the functions in W are continuous and strictly increasing with respect to their Kth coordinate, Assumption B.3 implies that

(T7.4) $\forall h \in W \quad Pr\{ (x^i, x^j) \mid h(x^i) = h(x) \} = 0$.

From (T7.1)-(T7.4) it then follows, by the arguments given in Matzkin (1990a, proof of Step 2), that

$\bar{h}^*$ uniquely maximizes $E[S_N(h)]$ over W. Q.E.D.

PROOF OF THEOREM 8: Let $r_{ij}(h)$ be defined as in the proof of Theorem 7. As it was shown in that proof,

(T8.1) $\forall h \in W \quad Pr\{ (x^i, x^j) \mid h(x^i) = h(x) \} = 0$.

Since convergence of $\{ h_k \} \subset W$ to $h \in W$ implies that

$\forall x \in X, \qquad h_k(x) \to h(x)$, it follows by the arguments given in Matzkin (1990a, Step 1, proof of (1.a.3)) that

(T8.3) $r_{ij}(h)$ is continuous on W a.s.

It then follows that

(T8.4) $E[S_N(h)]$ is continuous on W

(see Matzkin (1990a, Step 1, proof of (1.b.3)).

By the compactness of W (Theorem 1), (T8.2)-(T8.4), and the measurability of r_{ij} it follows that

(T8.5) $S_N(h)$ converges a.s. uniformly to $E[S_N(h)]$

(see Matzkin (1990a, proof of Step 3)).

The compactness of W, (T8.3), (T8.5), and Theorem 7 imply then that

$\hat{h}_N \to \bar{h}^*$ a.s.

(see Matzkin (1990a, proof of Step 4)). Q.E.D.

PROOF OF LEMMA 8: Let $h \in \tilde{W}$ be such that for some strictly increasing function $f: [\alpha,\gamma] \to [\alpha,\gamma]$, $h = f \cdot h^*$ and $h \neq h^*$; let $F = F^* \cdot f^{-1}$. Then, $(h,F) \in (\tilde{W} \times \Gamma)$ and for all $x \in X$

$$F^*(h^*(x)) = F \cdot f^{-1}(f \cdot h(x)) = F(h(x)).$$

Hence, $P(j|x;h^*,F^*) = P(j|x;h,F)$ for all $x \in X$ and for $j=1,2$. Thus, (h,F) is observationally equivalent to (h^*,F^*). Q.E.D.

PROOF OF LEMMA 9: Suppose that $(h,F) \in (W \times \Gamma)$ is observationally equivalent to (h^*,F^*). We show that then there exists a strictly increasing function f such that $h = f \cdot h^*$. Suppose that such a function f

does not exist. Then, there must exist some $(x_1,x_2) \in (X \times X)$ such that either

$[\ h(x_1) > h(x_2)$ and $h^*(x_1) \le h^*(x_2)\ ]$ or

$[\ h(x_1) \le h(x_2)$ and $h^*(x_1) < h^*(x_2)\ ]$.

Suppose w.l.o.g. that $h(x_1) > h(x_2)$ and $h^*(x_1) \le h^*(x_2)$. Then, since both h and h^* are continuous and strictly increasing with respect to the Kth coordinate, there exists $(x_1',x_2') \in (X \times X)$ such that

$$h(x_1') > h(x_2') \quad \text{and} \quad h^*(x_1') < h^*(x_2').$$

Our assumptions on F and F^* imply then that

$$F(h(x_1')) \ge F(h(x_2')) \quad \text{and} \quad F^*(h^*(x_1')) < F^*(h^*(x_2')).$$

Hence,

either $P(1|x_1';h,F) \ne P(1|x_1';h^*,F^*)$ or $P(1|x_2';h,F) \ne P(1|x_2';h^*,F^*)$.

But, either of these possibilities contradicts the hypothesis that (h,F) is observationally equivalent to (h^*,F^*). This contradiction shows that there exists a strictly increasing function f such that $h = f \cdot h^*$. By Lemma 2 h and h^* possess common least concave representatives. Lemma 4 and the definition of W imply then that $\bar{h} = \bar{h}^*$.

It remains to show that $\bar{F} = \bar{F}^*$. To show this, we note that, since (h,F) is observationally equivalent to (h^*,F^*), $F^*(h^*(x)) = F(h(x))$ for all $x \in X$. Hence, by the definitions of $\bar{F}$ and $\bar{F}^*$, $\bar{F}^*(\bar{h}^*(x)) = \bar{F}(\bar{h}(x))$ for all $x \in X$. Since $\bar{h} = \bar{h}^*$, this implies that $\bar{F}^* = \bar{F}$. Hence, $(\bar{h},\bar{F}) = (\bar{h}^*,\bar{F}^*)$.

Conversely, suppose that $(\bar{h},\bar{F}) = (\bar{h}^*,\bar{F}^*)$. Then, the definitions of $\bar{F}$ and $\bar{F}^*$ imply that

$$F^*(h^*(x)) = \bar{F}^*(\bar{h}^*(x)) = \bar{F}(\bar{h}(x)) = F(h(x)) \quad \text{for all } x \in X.$$

Hence, (h,F) is observationally equivalent to (h^*,F^*). Q.E.D.

PROOF OF THEOREM 10: Let $(h,F) \neq (\bar{h}^*,\bar{F}^*)$. Suppose first that $h \neq \bar{h}^*$. Then by Lemma 0, there exist $X_1 \subset X$ and $X_2 \subset X$ such that for all $x_1 \in X_1$ and $x_2 \in X_2$

(T10.1) $h(x_1) < h(x_2)$ and $\bar{h}^*(x_1) > \bar{h}^*(x_2)$.

Since, by Assumption C.2 and the definition of $\bar{F}^*$, $\bar{F}^*$ is strictly increasing and, by the definition of Γ, F is increasing, (T10.1) implies that for all $x_1 \in X_1$ and $x_2 \in X_2$,

(T10.2) $F(h(x_1)) \leq F(h(x_2))$ and $\bar{F}^*(\bar{h}^*(x_1)) > \bar{F}^*(\bar{h}^*(x_2))$.

It is then impossible that both

$$F(h(x_1)) = \bar{F}^*(\bar{h}^*(x_1)) \quad \text{and} \quad F(h(x_2)) = \bar{F}^*(\bar{h}^*(x_2)).$$

Suppose w.l.o.g. that $F(h(x_1)) \neq \bar{F}^*(\bar{h}^*(x_1))$ and

$$F(h(x_1)) < \bar{F}^*(\bar{h}^*(x_1)).$$

(The analysis is analogous when $F(h(x_1)) > \bar{F}^*(\bar{h}^*(x_1))$.) Since F and $\bar{F}^*$ are increasing and h and $\bar{h}^*$ are continuous on X, $F \cdot h$ and $\bar{F}^* \cdot \bar{h}^*$ are continuous at a.e. x_1. Hence, there exists $\delta > 0$ such that for a.e. $x \in N(x_1,\delta)$, $F(h(x)) < \bar{F}^*(\bar{h}^*(x))$. Since, by Assumption C.3, $N(x_1,\delta)$ possesses positive probability, it follows that

$$\int_{N(x,\delta)} g(x)\ F(h(x))\ dP_x(x) \neq \int_{N(x,\delta)} g(x)\ \bar{F}^*(\bar{h}^*(x))\ dP_x(x) .$$

Suppose now that $h = \bar{h}^*$ but $F \neq \bar{F}^*$. Then, since both F and $\bar{F}^*$ are monotone increasing on $[\alpha,\gamma]$, there exists a $t \in (\alpha, \gamma)$ such that t is a point of continuity of both F and $\bar{F}^*$ and $F(t) \neq \bar{F}^*(t)$. Let $x \in X$ be such that $\bar{h}^*(x) = t$. Since $\bar{h}^*$ is continuous at x, there exists a neighborhood N of x such that for all $x' \in N$

$F(h(x')) = F(\bar{h}^*(x')) < \bar{F}^*(\bar{h}^*(x'))$ if $F(t) < \bar{F}^*(t)$ and

$F(h(x')) = F(\bar{h}^*(x')) > \bar{F}^*(\bar{h}^*(x'))$ if $F(t) > \bar{F}^*(t)$.

Since, by Assumption C.3, N possesses positive probability,

$$\int_N g(x)\ F(h(x))\ dP(x) \neq \int_N g(x)\ \bar{F}^*(\bar{h}^*(x))\ dP(x) \ .$$

Q.E.D.

PROOF OF THEOREM 11: We will first show that our model satisfies the assumptions used in Lemmas 1-5 in Matzkin (1988). The result of the theorem will then follow from these lemmas, Theorems 1 and 10, and the compactness of Γ, by the arguments described in Matzkin (1988, proof of Theorem 3).

The definition of W and C.7 imply that Assumptions W.1 and W.2 in Matzkin (1988) are satisfied; and Assumptions Γ.1 - Γ.3 in Matzkin (1988) are implied by our assumptions on $\bar{F}^*$ and Γ . By Assumptions C.4, C.5, and C.6 in this paper, Assumptions G.1 - G.3 in Matzkin (1988) hold. Theorem 1 in this paper shows that Assumption W.5 in Matzkin (1988) is satisfied. Assumption W.6 in Matzkin (1988) holds from our Assumption C.5 and the continuity and strict monotonicity with respect to the K^{th} coordinate of the functions in W. Hence, our model satisfies Lemmas 1-5 in Matzkin (1988).

Since our set Γ is compact with respect to d_Γ (see Cosslett (1983)), it follows from Theorems 1 and 10 that

$$\lim_{N\to\infty} d(\hat{h}_N, \bar{h}^*) = 0 \quad \text{and} \quad \lim_{N\to\infty} d(\hat{F}_N, \bar{F}^*) = 0 \quad \text{a.s.}$$

(see Matzkin (1988), proof of Theorem 3). Q.E.D.

REFERENCES

BENVENISTE, L. M. and J.A. SCHEINKMAN (1979), " Differentiable Value Functions in Concave Dynamic Optimization Problems," Econometrica.

COSSLETT, S. (1983), "Distribution-Free Maximum Likelihood Estimation of the Binary Choice Model," Econometrica, 51, 765-782.

DEBREU, G. (1976), "Least Concave Utility Functions", Journal of Mathematical Economics, reprinted in G.Debreu, Mathematical Economics, Cambridge University Press, 1983.

FINETTI, B. de (1949), "Sulle Stratificazioni Convesse," Annali de Matematica Pura ed Applicata, 30, Serie 4, 173-183.

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

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

KANNAI, Y. (1977), " Concavifiability and Constructions of Concave Utility Functions," Journal of Mathematical Economics, 4, 1-56.

__________. (1980), " The ALEP Definition of Complementarity and Least Concave Utility Functions," Journal of Economic Theory, 22, 115-117.

KLEIN, R. W. and R. H. SPADY (1987), "Semiparametric Estimation of Discrete Choice Models," Bell Communications Research.

MADDALA G. S. (1983), Limited-Dependent and Qualitative Variables in Econometrics, Cambridge University Press, Cambridge.

MANSKI, C. (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.

__________. (1988), "Nonparametric and Distribution-free Estimation of the Threshold Crossing and Binary Choice Models," Cowles Foundation Discussion Paper No. 889, Yale University.

__________. (1990a), "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.

__________. (1990b), "Maximization of Discontinuous Functions over Nonparametric Sets of Concave Functions," Cowles Foundation, Yale University.

RAO, R.R. (1962), "Relation Between Weak and Uniform Convergence of Measures with Applications," Annals of Mathematical Statistics 33, 659-680.

NOTES

1. The Gaussian curvature of a C^2 function h at $x \in R^K$ is the determinant of the matrix

$$\begin{bmatrix} -D^2h(x) & Dh(x) \\ -Dh(x)^T & 0 \end{bmatrix} \frac{1}{\|Dh(x)\|^{K+1}} .$$

2. By $f \cdot g$ we mean the composition of the functions f and g.

3. A vector $\beta(h,x)$ is a *subgradient* of a concave function $h : X \rightarrow R$ at $x \in X$ if for all $y \in X$, $h(y) - h(x) \leq \beta(h,x)\ (y - x)$.

4. Note that the monotonicity of $G(\cdot)$ and the independence of η from x already imply that $P_{\eta|x_1}[\ G(h(x_1),\eta) \leq t] \geq P_{\eta|x_2}[\ G(h(x_2),\eta) \leq t]$.

5. Since the domain of the functions in Γ is bounded, d_Γ is topologically equivalent to the metric on the set of distribution functions that was employed by Cosslett (1983) and Matzkin (1988).

6. Kannai (1977,pp.52) notes that the least concave function constructed in his Theorem 2.4 is continuous on $K \times P_{conc,reg}$, where $P_{conc,reg}$ is the set of C^2 preference orderings for which the Gaussian curvature of the indifference surfaces $\{ x \mid x \sim y \}$ never vanishes and K is a compact set on which the preference orderings are defined.

7. A set $C \subset R^K$ is *monotone increasing* if for all $x \in C$ and all $e \in R^K$ such that $e_k \geq 0$ $(k=1,\ldots,K)$, $(x + e) \in C$.2