9-1-1988

Nonparametric and Distribution-Free Estimation of the Binary Choice and the Threshold-Crossing Models

Rosa L. Matzkin

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 889

NONPARAMETRIC AND DISTRIBUTION-FREE ESTIMATION OF THE BINARY CHOICE AND THE THRESHOLD-CROSSING MODELS

by

Rosa L. Matzkin

NONPARAMETRIC AND DISTRIBUTION-FREE ESTIMATION

OF THE BINARY CHOICE

AND THE THRESHOLD-CROSSING MODELS

by

Rosa L. Matzkin*
Cowles Foundation
Yale University

December 1987

Revised : September 1988

ABSTRACT

This paper studies the problem of nonparametric identification and estimation of binary threshold-crossing and binary choice models. First, conditions are given that guarantee the nonparametric identification of both the function of exogenous observable variables and the distribution of the random terms. Second, the identification results are employed to develop strongly consistent estimation methods that are nonparametric in both the function of observable exogenous variables and the distribution of the unobservable random variables. The estimators are obtained by maximizing a likelihood function over nonparametric sets of functions. A two-step constrained optimization procedure is devised to compute these estimators.

---

* The support of the National Science Foundation through Grants No. SES-8720596 and SES-8713532 is gratefully acknowledged. I would like to thank James Heckman for his encouragement and Chuck Manski for discussions about response models. For his comments, I would like to thank Vassilis Hajivassiliou.

1. INTRODUCTION

In recent years, there has been increasing interest in the study of binary threshold-crossing models and binary choice models. These models have been employed to study a wide variety of problems. The applications have included labor force participation, choice of education, and choice of mode of transportation. The analysis of these models has been originally parametric and most recently semiparametric. This paper introduces a fully nonparametric analysis.

In binary threshold-crossing models the value of a dichotomous observable variable y is 1 if the value of a function h plus the value of a random term η is above some threshold value; the value of y is 0 otherwise. The value of h depends on observable exogenous variables. For example, h may denote willingness to work as a function of socioeconomic characteristics, and η may denote a function of unobservable variables influencing the individual's willingness to work. The variable y then equals 1 when an individual participates in the labor force, and y equals zero otherwise.

In binary choice models, h is the difference between the values of a subutility, V, evaluated at the observable attributes of two alternatives. The value of η is the difference between the values of a subutility, e, evaluated at the unobservable attributes of the two alternatives. The observable variable y equals 1 when the value of h plus η is grater than or equal to zero. In a well known application of this model, the first alternative is commuting by bus and the second commuting by car. In this application, V is a function of cost, time, and socioeconomic characteristics, and e denotes the value of a function of unobservable attributes.

In the past, the estimation of threshold and discrete choice models has proceeded by specifying parametric structures for the function h and for the distribution F of η . Unlike the estimation of linear models, however, an erroneous specification of the distribution F may cause the estimator of the parameters of the correctly specified function h to be inconsistent. In pioneering work, Manski (1975,1985) proved that it is possible to estimate the parameters of h consistently without requiring the distribution of η to be parametric. Recently, many other distribution-free methods of estimating h [V] or (h,F) [(V,F)] have been developed. The maximum likelihood distribution-free estimator of Cosslett (1983) for binary choices, the maximum rank correlation estimator of Han (1987) for generalized regression models, Ichimura's (1986) estimator for single-index models, Klein and Spady's (1987) estimator for discrete-choice models, and Stoker's (1985) average-derivatives estimator are some of the new distribution-free estimators that apply to the limited dependent variable models studied in this paper. All these methods still rely on a parametric structure for h .

Matzkin (1987) presented a method of estimating a monotone and concave subutility function V in polychotomous choice models. This estimation method provided a strongly consistent estimator of the function V . The method did not require V to possess a parametric structure, but it did require the distribution of the random terms e to belong to a parametric family. The parameters of the distribution of e were also consistently estimated. Another estimation method in the semiparametric vein is the flexible form methods of Gallant (1981,1982).

In this paper, we introduce a nonparametric, distribution-free method of estimating the functions h and F in the threshold-crossing model and the functions V and F in the binary choice model, without requiring either h [V] or the distribution F of η to possess a parametric structure. This new method exploits the knowledge economists possess about properties of the function h [V] ; for example, h [V] may be known to belong to the set of concave, monotone, and linearly homogeneous functions. Instead of estimating h [V] from a parametric family of functions, the new methods obtain an estimator of h [V] from a subset W of functions possessing this particular set of properties, or some other set of properties that may be implied by economic theory. The function F is only assumed to belong to a set Γ of increasing functions whose range lies in the interval [0,1]. Since economic theory does not in general imply any properties for F , it is desirable to develop methods that do not require F to satisfy any specific conditions.

Section 2 defines formally the binary threshold-crossing and choice models. Section 3 presents a set of conditions on the set W of functions h [V] under which it is possible to identify (h,F) [(V,F)] within $(W \times \Gamma)$. This section also discusses the application of these identification results to develop various new nonparametric, distribution-free methods by combining and modifying existing semiparametric methods. Section 4 contains examples of sets of functions that possess properties typically assumed in microeconomic theory and satisfy the identification conditions presented in Section 3.

Section 5 presents a strongly consistent estimation method for (h,F) [(V,F)]. The method is based upon the combination and modification of

Cosslett's (1983) distribution-free estimator and Matzkin's (1987) semiparametric estimator. This new estimation method proceeds by maximizing a likelihood function over the set $(W \times \Gamma)$. In Section 6, we prove that this estimator is strongly consistent. The proof is based upon Wald's (1949) result about the strong consistency of the maximum likelihood estimator. Section 7 presents some brief concluding remarks and summarizes the main results of the paper.

2. THE MODELS

The binary threshold-crossing model and the binary choice model are described below. An extensive list of empirical applications of these models can be found in Maddala (1983).

The binary threshold-crossing model:

In this model, we assume that the value of an observable dichotomous variable y is determined by

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

In (1), $r \in T_t$ denotes a vector of observable exogenous variables, $T_t \subset R^K$, $h^* : T_t \to R$ is an unknown function, and η is an unobservable random variable. $1[\cdot]$ is a logical operator that equals 1 when $[\cdot]$ is true and equals zero otherwise. The random variable η is assumed to be independent of r and to possess an unknown cumulative distribution function $F^* : R \to [0,1]$. The vector r is assumed to possess an unknown probability density function g that induces a probability measure G .

The probability that $y = 1$ when the vector of observable variables is r will be denoted by $\Pr(y = 1|r)$. This probability depends only on the value that $h*$ attains at r and the value that $F*$ attains at $h*(r)$, since by (1), $\Pr(y = 1|r) = F*(h*(r))$.

The conditional log-likelihood of a sample of N independent observations $x^{(N)} = \{y^i, r^i\}_{i=1}^N$ when $h^* = h$ and $F^* = F$ is then

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

The binary choice model:

In this model, the value of the dependent observable variable, y, which equals one when the first of two alternatives is chosen and equals zero otherwise, is assumed to be determined by

$$(2) \qquad y = 1[V*(r_1) + e_1 \geq V*(r_2)) + e_2] .$$

In (2), $r_j \in T_d$ denotes a vector of observable attributes corresponding to alternative j $(j = 1, 2)$, $V* : T_d \to R$ is an unknown subutility function, and e_j is an unobservable random term representing the value of a subutility function on unobservable attributes corresponding to alternative j $(j = 1, 2)$. For each alternative j , the vector r_j may include the alternative, $j \in A$, socioeconomic characteristics of the consumer, $s \in S$, and attributes of the alternative, $r_j \in T_t$. Hence, $T_d \subset (A \times S \times Z)$. The vector (s, z_1, z_2) is assumed to possess a probability density, g, that induces a probability measure G .

The random vector (e_1, e_2) is assumed to be distributed independently of (s, z_1, z_2) and to possess an unknown probability density. Let η denote $-(e_1 - e_2)$. The random variable η is then distributed independently of (s, z_1, z_2) with an unknown cumulative distribution function $F^* : R \to [0,1]$, which is induced by the density of (e_1, e_2).

The probability that $y = 1$ given (r_1, r_2) is then

$$\Pr(y = 1 | r_1, r_2) = F^*(V^*(r_1) - V^*(r_2)) .$$

This is a binary threshold-crossing model with $h^*(r_1, r_2) = V^*(r_1) - V^*(r_2)$ and $\eta = e_2 - e_1$.

The conditional log-likelihood of a sample of N independent observations $x^{(N)} = \{y^i, r_1^i, r_2^i\}_{i=1}^N$ when $V^* = V$ and $F^* = F$ is then

$$\mathcal{L}(x^N, V, F) = \sum_{i=1}^{N} \{y^i \log(F(V(r_1^i) - V(r_2^i)) + (1-y^i)\log(1 - F(V(r_1^i) - V(r_2^i)))\} .$$

3. IDENTIFICATION

In this section, we study the identification of nonparametric functions (h^*, F^*) and (V^*, F^*), within sets of nonparametric functions, in the models described in Section 2.[1]

We say that a pair of functions (h^*, F^*) $[(V^*, F^*)]$ is identified in a set $(W \times \Gamma)$ if h^* belongs to W, F^* belongs to Γ, and any other pair (h,F) $[(V,F)]$ that belongs to $(W \times \Gamma)$ and is different from (h^*, F^*) $[(V^*, F^*)]$, induces a different probability density on the observable dependent variable. We say that two distribution functions F and F' are different if they attain different values on a set of positive Lebesgue

[1] The identification of binary threshold-crossing and binary choice models in which h^* belongs to a set of linear-in-parameters functions and F^* belongs to a set of nonparametric functions has been analyzed in Manski (1986).

measure with respect to η , and we say that two functions h and h' [V and V'] are different if they attain different values on a set of positive probability measure with respect to r . The formal definition of identification follows.

DEFINITION: *The pair* (h^*, F^*) $[\ (V', F')\ ]$ *is* *__identified__* *in the set* $(W \times \Gamma)$ *if for any pair* (h,F) $[\ (V,F)\ ]$ *in* $(W \times \Gamma)$ *such that* $(h,F) \neq (h^*, F^*)$ $[\ (V,F) \neq (V^*, F^*)\ ]$ *there exists a set* D *in the support of the probability density* g *such that for some value* y' *of* y *and all* $r \in D$

$$Pr(y = y'|r;\ h,\ F) \neq Pr(y = y'|r;\ h^*,\ F^*)$$

[respectively, $Pr(y = y'|r;\ V,\ F) \neq Pr(y = y'|r;\ V^*,\ F^*)$ *].*

In the models that have been described in Section 2, all we can observe from the data are the frequencies with which $y = 1$ for given values of r in the support of g . These frequencies are determined by the composition of two functions. To identify (h^*, F^*) in a set $(W \times \Gamma)$, this set needs to satisfy conditions that will allow us to separate the influence of F^* from the influence of h^* . Similarly, in binary choice models, to identify (V^*, F^*) in a set $(W \times \Gamma)$ we need to separate the influence of F^* from the influence of the difference between the value of V^* at two points. The following set of assumptions allows such a separation.

<u>Assumptions on the set of functions W :</u>

W.1. In the threshold-crossing model, W is a set of real-valued, continuous functions with domain T_t.

W.1'. In the binary choice model, W is a set of real-valued, continuous functions with domain T_d.

W.2. In the binary threshold-crossing model, $h^* \in W$.

W.2'. In the binary choice model, $V^* \in W$.

W.3. In the threshold-crossing model,

there exists a subset $\bar{T}$ of T_t such that

(i) for all $h, h' \in W$ and all $r \in \bar{T}$

$$h(r) = h'(r) \text{ , and}$$

(ii) for all $h \in W$ and all t in the domain of F^*,

there exists $r \in \bar{T}$ such that

$$h(r) = t \text{ .}$$

W.3'. In the binary choice model,

there exists a subset $\bar{T}$ of $S \times Z$ such that

(i) for all $V, V' \in W$ and all $(s, z_2) \in T$

$$V(2, s, z_2) = V'(2, s, z_2) \text{ , and}$$

(ii) for all $V \in W$ and all t in the domain of F^*,

there exists $(s, z_2) \in \bar{T}$ such that

$$V(2, s, z_2) = t \text{ .}$$

W.4. In the binary choice model,

(i) if for $j = 1, 2$, $r_j = (j, s, z_j) \in A \times S \times Z$, then for some $\bar{z} \in Z$, all $s \in S$, and all $V \in W$, $V(1,s,\bar{z}) = 0$;

(ii) if for $j = 1, 2$, $r_j = (j, z_j) \in A \times Z$, then for some $\bar{z} \in Z$, and all $V \in W$, $V(1,\bar{z}) = 0$;

(iii) if for $j = 1, 2$, $r_j \in S \times Z$, then for some $\bar{z} \in Z$, all $s \in S$, and all $V \in W$, $V(s,\bar{z}) = 0$;

(iv) if for $j = 1, 2$, $z_j \in Z$, then for some $\bar{z} \in Z$ and all $V \in W$, $V(\bar{z}) = 0$.

Section 4 presents several examples of sets, W, of functions satisfying these assumptions.

Assumption W.1 [W.1′] guarantees that all points in the domain of the functions in W will be observed. The continuity assumptions on the functions in W also will be employed to prove the consistency of our estimators. Note that when the support T_t is a finite number of points, the continuity assumption is trivially satisfied.

Assumption W.3 [W.3′] is critical for our results. Assumption W.3(i) [W.3′(i)] allows us to separate the influence of F^* from the influence of h^* (or from the difference between the value of V^* at two points) on the probability distribution of y. Since all functions in W attain the same values at $\bar{T}$, a difference in the probability of y given $r \in \bar{T}$ and (h,F) can only be accounted for by a difference between F and F^*. Similarly, in the binary choice model, a difference in the probability of y given $(z_1, z_2) \in \{(1,s,\bar{z}) \times \bar{T}$ and (V,F) can only be accounted for by a difference between F and F^*. Assumption W.3(ii) [W.3′(ii)] guarantees that the range of h over $r \in \bar{T}$ (or the range of $V^*(1,s,\bar{z}) - V^*(2, s, z_2)$ over $(s, z_2) \in \bar{T}$) contains the support of F^*. This assumption insures the identification of F^* from any F such that $F \neq F^*$. Since, if $F(t) \neq F^*(t)$, by Assumption W.3(ii) [W.3′(ii)] there exists $r^* \in \bar{T}$ (or $(s^*, z_2^*) \in \bar{T}$) such that for all $h \in W$, $h^*(r^*) = h(r^*) = t$ (or, for all $V \in W$, $V^*(1, s^*, \bar{z}) - V^*(2, s^*, z_2^*) = V(1, s^*, \bar{z}) - V^*(2, s^*, z_2^*) = t$). Hence, $F^*(h^*(r^*)) \neq F(h(r^*))$ (or, $F^*(V^*(1, s^*, \bar{z}) - V^*(2, s^*, z_2^*)) \neq F(V(1, s^*, \bar{z}) - V(2, s^*, z_2^*))$).

Note that in binary choice models in which V^* is a function of the alternatives, the set $\bar{T}$ need only exist in the domain of $V^*(2,\cdot)$. Note also that in the binary choice model, the set $\bar{T}$ may be a subset of $\{\bar{s}\} \times Z$, for some $\bar{s} \in S$, or a subset of $S \times \bar{r}$, for some $\bar{r} \in Z$; in other words, the functions in W may be assumed to attain the same values only for a particular value of socioeconomic characteristics or only for a particular value of attributes of the alternatives.

Assumption W.4 is only required in the binary choice model; it guarantees the identification of (V^*, F^*) from any (V, F^*) such that $V \neq V^*$. Assumption W.4(i) is analogous to the identification assumptions made about V^* , when V^* is assumed to be linear in parameters and to depend on socioeconomic characteristics of the consumer and alternative specific constants. Assumption W.4(ii) is the same as assumption W.4(i), except that it applies to the case in which V^* does not depend on socioeconomic characteristics. Assumptions W.4(iii) and W.4(iv) apply when the value of V^* at the observable attributes or characteristics does not depend on the alternatives. In this case, a consumer will attain the same V-utility value at any two alternatives that possess the same attributes. The meaning of Assumptions W.4(iii) and W.4(iv) is that all functions in W will attain the same value at some vector of attributes $\bar{z} \in Z$. This normalization is necessary because the probability of y given (z_1, z_2) depends only on the difference of V^* between two alternatives.

<u>Assumptions on the set Γ</u> :

Γ.1. Γ is the set of all monotone increasing functions on R with values in $[0,1]$.

Γ.2. $F^* \in \Gamma$.

Γ.3. F^* is strictly increasing.

Assumptions Γ.1 and Γ.2 guarantee the compactness of the set Γ to which F^* belongs. This property will be important in proving the strong consistency of our estimators. Assumption Γ.3 will allow us to identify (h, F^*) from (h^*, F^*) when $h \neq h^*$ or, in the binary choice model, (V, F^*) from (V^*, F^*) when $V \neq V^*$. In the threshold-crossing model, this identification property holds because if $h(r) \neq h^*(r)$ the strict monotonicity of F^* implies that $F^*(h(r)) \neq F^*(h^*(r))$. In the binary choice model, this holds because when $V(2, s, z_2) \neq V^*(2, s, z_2)$ for some $(s, z_2) \in S \times Z$, Assumption W.4 and the strict monotonicity of F^* will imply that $F^*(V(1, s, \bar{z}) - V(2, s, z_2)) \neq F^*(V^*(1, s, \bar{z}) - V(2, s, z_2))$. If, on the other hand, $V(2, s, z_2) = V^*(2, s, z_2)$ for all $(s, z_2) \in S \times Z$, it must then be that $V(1, s, z_1) \neq V^*(1, s, z_1)$ for some $(s, z_1) \in S \times Z$. In this case it follows that for any $z_2 \in Z$ the strict monotonicity of F^* implies that $F^*(V(1, s, z_1) - V(2, s, z_2)) \neq F^*(V^*(1, s, z_1) - V^*(2, s, z_2))$.

Assumptions Γ.1-Γ.3 are also made in Cosslett's (1983) semiparametric distribution-free method.

<u>Assumptions on the probability density g</u> :

G.1. In the binary threshold-crossing model, the support of the probability measure G of the vector r of exogenous observable variables is T_t , the domain of h^* .

G.1'.In the binary choice model, the support of the probability measure G of (s, z_1, z_2) is $S \times Z^2$.

G.2. In the binary threshold-crossing model, at least one coordinate of r possesses a Lebesgue density conditional on the other components of r.

G.2'.In the binary choice model, either

(i) there exists a coordinate s_k of $s \in S$ such that the probability density of s_k conditional on the remaining coordinates of $(s, z_1, z_2) \in S \times Z^2$ is a Lebesgue density, or

(ii) there exists a coordinate $z_{k'}$ of $z \in Z$ such that the probability density of $(z_{1,k'}, z_{2,k'})$, conditional on the remaining coordinates of $(s, z_1, z_2) \in S \times Z^2$, is a Lebesgue density.

Assumption G.1 [G.1'] guarantees that the elements in the domain of the functions in W can be observed with positive probability. If this assumption was not satisfied, we could not distinguish h^* [V^*] from an h [V] that differs from h^* [V^*] only at the unobserved points. Assumption G.2 [G.2'] is made in all semiparametric distribution-free methods of estimating binary threshold-crossing and choice models; it guarantees that all points in the domain of F^* will be attained with positive probability.

Our main results in this section are the following theorems.

Identification in the threshold-crossing model:

THEOREM 1: *Suppose that the binary threshold-crossing model satisfies assumptions W.1-W.3, Γ.1-Γ.3 and G.1-G.2. Then,* (h^*, F^*) *is identified in* $(W \times \Gamma)$.

Identification in the binary choice model:

THEOREM 2: *Suppose that the binary choice model satisfies assumptions W.1′, W.3′, W.4, Γ.1-Γ.3 and G.1′-G.2′. Then,* (V^*, F^*) *is identified in* $(W \times \Gamma)$.

The proofs of Theorems 1 and 2 are presented in Appendix A. Examples of sets of functions W that satisfy our identification conditions are discussed in the next section.

The identification results of the theorems above can be employed to develop strongly consistent estimation methods for (h^*, F^*) $[(V^*, F^*)]$ or for h^* $[V^*]$. In particular, existing semiparametric methods can be combined and modified to be nonparametric in both h^* $[V^*]$ and F^* and still provide consistent estimators. This requires, as a first step, to restrict h^* $[V^*]$ to belong to a subset of a set of functions W satisfying the Assumptions W.1-W.3 [W.1′- W.3′, and W.4] stated above. As a second step, additional restrictions on Γ and W are imposed to insure the strong consistency of the estimation method. In Section 5, we describe one such estimator and in Section 6, we prove its strong consistency. This estimator is based upon Cosslett's (1983) semiparametric distribution-free method and Matzkin's (1987) method. An example that demonstrates how the basic identification results obtained in this paper can be applied to develop a strongly consistent estimator of a function h^* of observable exogenous variables in generalized regression models is given in Matzkin (1988).

4. EXAMPLES

In this section, we present several examples of sets of functions W that satisfy the assumptions made about W in the previous section.

EXAMPLE 1: In this example, we consider sets W of *homogeneous of degree one functions*,[2] for the threshold-crossing model.

Suppose that $T_t = R_+^K$ and that the domain of F^* is bounded below by 0. Let $r^* \in \text{int}(T_t)$ and $\alpha \in R$ be given, and let W be the set of all continuous and homogeneous of degree one functions $h : T_t \to R$ such that $h(r^*) = \alpha$. Assume that r is distributed with a Lebesgue density with support T_t . Then, W satisfies Assumptions W.1 and W.3. The set $\bar{T}$ of Assumption W.3 is the ray $\{r \in T_t | r = \lambda r^* \text{ for some } \lambda \in R_+\}$. Since for all $h \in W$ and all r in this set $\bar{T}$,

$$(1.1) \qquad h(r) = h(\lambda r^*) = \lambda h(r^*) = \lambda\alpha$$

[2]A function $h : T \to R$ is *homogeneous of degree one* if, for all $r \in T$ and all $\lambda \geq 0$, $h(\lambda r) = \lambda h(r)$.

for $\lambda \in R$ for which $r = \lambda r^*$, W satisfies W.3(i). Since all functions in W map $\bar{T}$ onto $[0,\infty)$, W satisfies W.3(ii).

The assumption that W is a set of homogeneous of degree one functions implies that each function in W can be determined by one level set and that the value of the function increases linearly across level sets. The rate of increase in the value of the functions in W is determined by the specified value α at z^* . Different functions in W attain the same values at r^* and at the ray $\bar{T}$ that passes through r^* , but differ in their level sets. (See Figure 1.)

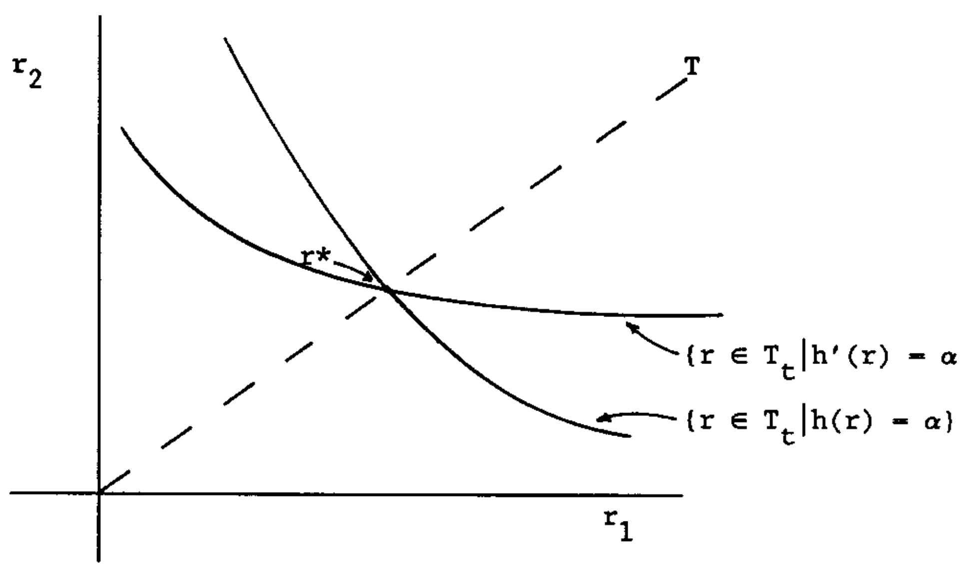

FIGURE 1

Semiparametric, distribution-free methods of estimating threshold-crossing models typically assume that $h^*(r) = r\,\theta^*$ for some parameter θ^* which belongs to the set

$$(1.2) \qquad \{\theta \in R^K \mid \|\theta\| = 1\} .$$

Fixing the norm of the parameter values is analogous to our assumption that $h(r^*) = \alpha$; both fix the rate of increase of the functions. Linearity of the functions in r , however, implies not only that the functions are homogeneous of degree one, but also that their level sets are hyperplanes. (See Figure 2.)

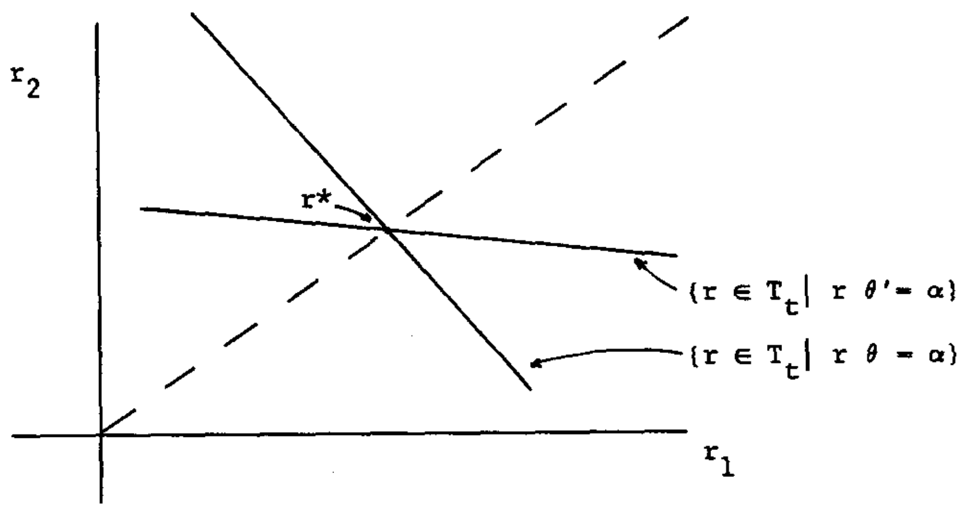

FIGURE 2

EXAMPLE 2: In this example, we consider sets of functions that are *additively separable into the value of one variable and any continuous function of the remaining variables.* This set will allow us to identify (h^*, F^*) in the threshold-crossing model.

Let E be a subset of R^{K-1}, and let W be the set of all functions $h : R \times E \to R$ for which there exists a function $t : E \to R$ such that

(2.1) $t(0) = 0$, and

(2.2) $h(r_1, r_2, \ldots, r_K) = r_1 + t(r_2, \ldots, r_K)$ for all $r \in R^K$.

Assume that $0 \in E$; that r_1 is distributed, conditional on $(r_2, \ldots, r_K)$, with a continuous distribution; and that the support of G is $R \times E$. Then W satisfies Assumptions W.1 and W.3 for the threshold-crossing model. The set $\bar{T}$ of Assumption W.3 is the ray $\{r \in R^K | r_k = 0,\ k = 2, \ldots, K\}$. By (2.1) and (2.2), all functions in W attain the same values at any elements of this set $\bar{T}$. Moreover, they map $\bar{T}$ onto R and hence, this W set satisfies Assumption W.3(ii). By the assumptions made about G, $\bar{T}$ is included in the support of G.

Note that the conditions on this set of functions are very weak. Apart from the additive separability in (2.2), the normalization condition (2.1), and the continuity of the t functions, no additional restrictions need to be imposed in W to identify (h^*, F^*). If it is desirable, in some particular application, to impose additional properties on the functions in W, Theorem 1 will still hold.

EXAMPLE 3: We consider in this example sets of functions that are *additively separable into a homogeneous of degree one function and a continuous function.* We will show how the conditions in W.4 can be imposed on these sets to identify (V^*, F^*) in the binary choice model.

Assume that the vector $z \in Z$ has a continuous distribution, conditional on s. Let $\bar{s} \in S$, $\bar{z} \in Z$, and $\alpha \in R$ be given. Let W be the set of all continuous functions on $A \times S \times Z$ such that for all V in W there exist functions $v : A \times S \to R$ and $w : Z \to R$ satisfying

(3.1) $\quad v(1,s) = 0$ for all $s \in S$, (3.2) $v(2,\bar{s}) = 0$,

(3.3) w is homogeneous of degree one,

(3.4) $w(\bar{z}) = 0$, and

(3.5) for all $(j,s,z) \in A \times S \times Z$, $V(j,s,z) = v(j,s) + w(z)$.

If the support of G is $S \times Z^2$ and $V^* \in W$, the set W satisfies Assumptions W.1′- W.3′, and W.4. By (3.1) and (3.4), W satisfies Assumption W.4. By (3.2)-(3.4) and arguments similar to those given in Examples 1 and 2, W satisfies Assumption W.3′ with $\bar{T} = \{(s, z_2) | s = \bar{s}, z_2 = \lambda\bar{z}$ for some $\lambda \geq 0\}$.

This particular example is widely applicable because the functions v of socioeconomic characteristics are not required to possess any particular properties other than continuity. In applications in which economic theory has no particular implications about these functions, this is a very useful feature.

EXAMPLE 4: We consider in this example an alternative *set of additively separable functions* for the binary choice model.

For any $s \in S$, let $s_{-k} = (s_1, s_2, \ldots, s_{k-1}, s_{k+1}, \ldots)$. Assume that the coordinate s_k of $s \in S$ possesses a continuous distribution, conditional on (s_{-k}, z_1, z_2) .

Let $\bar{s} \in S$, $\bar{z} \in Z$, and $\alpha \in R$ be given. We will let W be the set of all continuous functions on $A \times S \times Z$ such that for all $V \in W$ there exist functions $t : A \times S \to R$ and $w : Z \to R$ satisfying

(4.1) $s_k + t(1, s_{-k}) = 0$ for all $s \in S$, (4.2) $t(2, \bar{s}_{-k}) = \alpha$,

(4.3) $w(\bar{z}) = 0$, and

(4.4) for all $(j,s,z) \in A \times S \times Z$, $V(j,s,z) = s_k + t(j, s_{-k}) + w(z)$.

Assume that the support of G is $S \times Z^2$, and that $V^* \in W$. Then, by following arguments similar to those given for the sets of functions W in Examples 1, 2, and 3, it is easy to show that any set of functions W satisfying the conditions (4.1)-(4.4) satisfies Assumptions W.1′, W.3′, and W.4.

5. ESTIMATION

In this section, we present a particular method of estimating the pair of functions (h^*, F^*) in the binary threshold-crossing model and the pair (V^*, F^*) in the binary choice model. The method proceeds by finding the pair (h,F) $[(V,F)]$ that maximizes the likelihood of the observations over the set $(W \times \Gamma)$.

Since for any finite number of observations, the value of the likelihood $\mathcal{L}$ at any function $h \in W$ $[V \in W]$ depends on h $[V]$ only through the value that h $[V]$ attains at the vectors $r^1, \ldots, r^N$ $[(1, s^1, z_1^1), (2, s^1, z_2^1), \ldots, (1, s^N, z_1^N), (2, s^N, z_2^N)]$, the maximization of $\mathcal{L}$ over W can be transformed into the maximization of $\mathcal{L}$ over the set of all finite dimensional vectors $(h^1, \ldots, h^N)$ $[(V_1^1, V_2^1, \ldots, V_1^N, V_2^N)]$, for which there exists a function $h \in W$ $[V \in W]$ with $h(r^i) = h^i$ for $i = 1, \ldots, N$ $(V(j, s^i, z_j^i) = V_j^i$ for

$i = 1, \ldots, N$; $j = 1, 2)$. Similarly, since the value of the likelihood function $\mathcal{L}$ at any $F \in \Gamma$ depends on F only through the value that F attains at $h^1, \ldots, h^N$ [$V_1^1, \ldots, V_2^N$] , the maximization of $\mathcal{L}$ over Γ can be transformed into the maximization of $\mathcal{L}$ over the set of all vectors $F^1, \ldots, F^N$ [$F_1^1, F_2^1, \ldots, F_1^N, F_2^N$] for which there exists a function $F \in \Gamma$ with $F(h_i) = F^i$, $i = 1, \ldots, N$ [$F(V_j^i) = F_j^i$, $i = 1, \ldots, N$; $j = 1, 2$]. Cosslett (1983) characterized the set of vectors $F^1, \ldots, F^N$ for which there exists a function $F \in \Gamma$ with $F(h^i) = F^i$ $(i = 1, \ldots, N)$, and Matzkin (1987) characterized the set of vectors $h^1, \ldots, h^N$ [$V_1^1, \ldots, V_2^N$] for which there exists a concave and monotone function h [V] with $h^i = h(r^i)$, $i = 1, \ldots, N$ [$V_j^i = V(r_j^i)$, $i = 1, \ldots, N$; $j = 1, 2$]. We next employ the above characterizations and demonstrate how to estimate nonparametrically a binary threshold-crossing model in which the set W satisfies the conditions in Example 1 and a binary choice model in which the set W satisfies the conditions in Example 4.

Let us consider first the binary threshold-crossing model. We will assume that h^* belongs to the set W of functions $h : T_t \to R$ that satisfy the conditions in Example 1 of Section 4 and that, in addition, are concave, monotone increasing, and possess subgradients bounded by $B \in R_+^K$. From the theory of optimization it follows that these or similar properties are satisfied by any indirect utility function or profit function.

Suppose that N independent observations $\{y^i, r^i\}_{i=1}^N$ are given, where $r^i \in T_t \subset R^K$. For each $i = 1, \ldots, N$ we denote $h(r^i)$ by h^i and $F(h^i)$ by F^i . Then, the estimation method proceeds by finding the values of $h^i \in R$, $T^i \in R^K$ $(i = 0, 1, 2, \ldots, N, N+1)$, and $F^i \in R$ $(i = 1, \ldots, N)$ that maximize

(5.1) $$\sum_{i=1}^{N} \{y^i \log F^i + (1 - y^i)\log(1 - F^i)\}$$

subject to

(5.2) $$h^0 = 0 \ , \quad h^{N+1} = \alpha \ , \quad r^0 = 0 \ , \quad r^{N+1} = r^* \ ,$$

(5.3) $$h^i \leq T^r r^i \ , \quad i, r = 0, 1, \ldots, N, N+1$$

(5.4) $$h^i = T^i r^i \ , \quad i = 0, 1, \ldots, N, N+1$$

(5.5) $$0 \leq T^i \leq B \ , \quad i = 0, 1, \ldots, N, N+1$$

(5.6) $$0 \leq F^i \leq 1 \ , \quad i = 1, \ldots, N,$$

(5.7) $$(F^i - F^r) \cdot (h^i - h^r) \geq 0 \ , \quad i, r = 1, \ldots, N.$$

The constraint in (5.2) characterizes vectors $(h^0, \ldots, h^{N+1})$ for which there exists $h : T_t \rightarrow R$ with $h(0) = 0$ and $h(r^*) = \alpha$.

Constraints (5.3), (5.4), and (5.5) are in the spirit of Afriat's (1972) inequalities. In Lemma C.1 in Appendix C, we show that they characterize vectors for which there exists a monotone, concave, homogeneous of degree one function $h : T_t \rightarrow R$ whose subgradients are bounded by B and which satisfies for $i = 0, \ldots, N+1$, $h(r^i) = h^i$. The constraints in (5.6) and (5.7) characterize vectors $F^1, \ldots, F^N$ for which there exists $F \in \Gamma$ with $F^i = F(h^i)$, $i = 1, \ldots, N$.

Suppose that $\hat{h}^0, \ldots, \hat{h}^{N+1}$, $\hat{T}^0, \ldots, \hat{T}^{N+1}$, $\hat{F}^1, \ldots, \hat{F}^N$ is a solution of the above maximization problem. Then, a particular estimate of (h*, F*) can be obtained by interpolating between these obtained values. For example, we may interpolate linearly between $\hat{F}^1, \ldots, \hat{F}^N$ and employ the following interpolation for $\hat{h}^0, \ldots, \hat{h}^{N+1}$, $\hat{T}^0, \ldots, \hat{T}^{N+1}$:

$$\hat{h}(r) = \min \{T^i r | i = 0, 1, \ldots, N, N+1 \} .$$

This interpolation is based upon Afriat (1972).

Following Cosslett (1983), we propose to solve the constrained maximization problem presented in (5.1)-(5.7) in two steps. The first step proceeds by finding, for any given vector $(h^0, \ldots, h^{N+1})$, the vector $\bar{F}^1, \ldots, \bar{F}^N$ that solves the following problem:

$$\text{Maximize } \sum_{i=1}^{N} \{y^i \log F^i + (1 - y^i)\log(1 - F^i)\}$$

(5.8)

$$\text{subject to } 0 \le F^i \le 1 , \quad i = 1, \ldots, N$$

$$(F^i - F^r)(h^i - h^r) \ge 0 , \quad i, r = 1, \ldots, N .$$

This can be done by employing an algorithm introduced by Asher <u>et al</u>. (1955). The optimal value of the objective function of (5.8) that is obtained from this algorithm depends only on $h^0, \ldots, h^{N+1}$. In the second step, the optimal value of the objective function is maximized over all vectors $(h^0, \ldots, h^{N+1})$ that satisfy (5.2)-(5.5) for some $T^0, \ldots, T^N \in R^K$. The algorithm of Asher <u>et al</u>. (1955) is described in Cosslett (1983, p. 773); its main steps are repeated in Appendix C.

For particular applications, some of the properties assumed about the functions in W can either be eliminated or changed. This would only require eliminating or changing some of the constraints in (5.2)-(5.5). For example, the functions in W may be assumed to be convex instead of concave; this would only require reversing the inequality sign in (5.3). They may not be required to be monotone, in which case 0 may be substituted by $-B$ in (5.5). In addition, they may not be required to possess uniformly bounded subgradients, in which case the second inequality in (5.5) is eliminated.

Let us consider now a binary choice model. We will assume that V^* belongs to the set W of functions $V : A \times S \times Z \to R$ that satisfy the conditions in Example 4 of Section 4. We will also assume that the functions $w : Z \to R$ are concave, $Z \subset R^K$, the vectors z_1 and z_2 are continuously distributed conditional on s, the vector s_{-k} has a discrete distribution conditional on (s_{-k}, z_1, z_2) with Q points of support, $\bar{s}_{-k}$ is one of the Q points, and that the values of the functions t are bounded by $B_1 \in R$ and $B_2 \in R$ $(\alpha \in [B_1, B_2])$.

Let $\{y^i, s^i, r_1^i, r_2^i\}_{i=1}^N$ be N independent observations. For each i and functions t, w, and F, we will denote $t(2, s_{-k}^i)$ by t^i, $w(z_j^i)$ by w_j^i, and $F(w_1^i - s_k^i - t^i - w_2^i)$ by F^i. Then the estimation proceeds by finding values t^i, w_j^i, and F^i for $i = 1, \ldots, N$, $j=1, 2$, subgradients $T_j^i \in R^K$ for $i = 1, \ldots, N$, $j = 1, 2$, and a subgradient $T^0 \in R^K$ that maximize

$$(5.9) \qquad \sum_{i=1}^{N} \{y^i \log F^i + (1 - y^i)\log(1 - F^i)\}$$

subject to

$$(5.10) \qquad s^0 = \bar{s}_{-k}, \quad t^0 = \alpha$$

(5.11) $$t^i = t^v \quad \text{if} \quad s^i = s^v, \qquad i, v = 0, \ldots, N$$

(5.12) $$B_1 \le t^i \le B_2 \;, \quad i = 1, \ldots, N$$

(5.13) $$w_j^i \le w_c^v + T_c^v(z_j^i - z_c^v) \;, \quad i, v = 1, \ldots, N \;; \quad j,c = 1, 2$$

(5.14) $$w_j^i \le T^0 z_j^i \;, \qquad i = 1, \ldots, N \;; \quad j = 1, 2$$

(5.15) $$0 \le w_c^v - T_c^v z_c^v \;, \quad v = 1, \ldots, N \;; \quad c = 1, 2$$

(5.16) $$0 \le F^i \le 1 \;, \quad i = 1, \ldots, N$$

(5.17) $$(F^i - F^v)((w_1^i - s_k^i - t^i - w_2^i) - (w_1^v - s_k^v - t^v - w_2^v)) \ge 0 \;, \quad i, v = 1, \ldots, N \;.$$

Constraints (5.10)-(5.12) correspond to our assumptions on the t functions, constraints (5.13)-(5.15) correspond to our assumptions on the w functions, and constraints (5.16) and (5.17) correspond to our assumptions on the F functions. A constraint of the form (5.11) needs to be imposed when it is likely that we observe repeatedly a point in the domain of a function.

The calculation of the estimator proceeds in two steps, similarly to the estimation of the threshold-crossing model described above.

The maximum likelihood estimation method that has been just described involves solving a large constrained optimization problem and maximizing a discontinuous function. In simulation experiments, exterior penalty methods (cf. Fiacco and McCormick (1968)) proved to be a fruitful approach to solve these kinds of problems.

In the next section, we show that the maximum likelihood estimation method described in this section is strongly consistent.

6. CONSISTENCY

In this section, we show that the estimators of (h^*, F^*) and of (V^*, F^*), obtained by maximizing the likelihood of the observations over $(W \times \Gamma)$, are strongly consistent. This result requires adding some assumptions to the identification assumptions stated in Section 3.

The strong consistency of the estimator of F^* is obtained with respect to the metric $d_\Gamma : \Gamma \times \Gamma \to R$ defined by

$$(6.1) \qquad d_\Gamma(F, F') = \int |F(t) - F'(t)| e^{-\|t\|} dt ,$$

where the integration is with respect to the Lebesgue measure and over the support of F^*. The strong consistency of the estimator of h^* $[V^*]$ is obtained with respect to two different metrics, d_w and d'_w. The metric $d_w : W \times W \to R$ is defined by

$$(6.2) \qquad d_w(h,h') = \int |h(r) - h'(r)| e^{-\|r\|} dG(r)$$

in the threshold-crossing model, and by

$$(6.3) \quad d_w(V,V') = \sum_{j=1}^{2} \int |V(j, s, z_j) - V'(j, s, z_j)| e^{-\|(s,z_j)\|} dG(s, z_1, z_2)$$

in the binary choice model. The alternative metric $d'_W : W \times W \to R$ is defined by

$$(6.4) \quad d'_W(h, h') = \inf \{ t | \; G(\{r \in T_t | \; |h(r) - h'(r)| > t\}) = 0 \}$$

for the threshold-crossing model and by

$$(6.5) \quad d'_W(V, V') = \inf \{ t | \; G(\{(s, z_1, z_2) | |V(j, s, z_j) - V'(j, s, z_j)| > t \text{ for some } j \in \{1,2\}\}) = 0 \}$$

for the binary choice model. Hence d_W is a modified $\mathcal{L}^1$ norm and d'_W is the essential supremum norm with respect to the probability measure G .

We next present the set of additional assumptions on W and G that will be made in Theorems 3, 4, 5, and 6.

<u>Assumptions on the set of functions W</u> :

W.5. The set W is compact with respect to the metric d_W .

W.5′. The set W is compact with respect to the metric d'_W .

W.6. For all functions h in W , the value of h possesses an absolutely continuous distribution.

W.6′. For all functions V in W , the value of V possesses an absolutely continuous distribution.

W.7. The functions in W are monotone increasing.

The compactness assumptions substitute for the usual assumption that the probability density of the observations tends to zero as the norm of the parameters tends to infinity. Compactness is also employed to prove the measurability of the supremum of the probability densities over neighborhoods in the space $(W \times \Gamma)$. Assumption W.6 [W.6′] is needed to prove the almost sure continuity of the probability density of the observations in the pair (h,F) [(V,F)]. Assumption W.7 is only employed when the consistency results are with respect to the metric d_W. This assumption, together with Assumption W.1, guarantees that convergence with respect to d_W implies pointwise convergence. This pointwise convergence is employed to prove the continuity of the probability density of the observations in the pairs (h,F) [(V,F)] and the measurability of probability densities over neighborhoods in the space $(W \times \Gamma)$.

Assumptions on the probability density g :

G.3. In the binary threshold-crossing model,

$$\int |\log g(r)|\ g(r)\ dr < \infty \ .$$

G.3′.In the binary choice model,

$$\int |\log g(s,z_1,z_2)|\ g(s,z_1,z_2)\ d(s,z_1,z_2) < \infty \ .$$

Assumption G.3 [G.3′] is needed to prove the integrability of several functions of the probability density of the observations. Many probability densities satisfy this assumption; in particular, any bounded density whose support is compact satisfies Assumption G.3 [G.3′].

The consistency results are stated in Theorems 3 and 4 for the threshold-crossing model and in Theorems 5 and 6 for the binary choice model. Theorems 3 and 5 concern the case in which the convergence of $\hat{h}_N$ [$\hat{V}_N$] to h^* [V^*] is with respect to the metric d_W . Theorems 4 and 6 state the consistency results for the case in which the convergence of $\hat{h}_N$ [$\hat{V}_N$] to h* [V*] is with respect to the metric d'_w . We next state the theorems.

THEOREM 3: *Suppose that the binary threshold-crossing model satisfies Assumptions W.1-W.3, W.5-W.7, Γ.1-Γ-3, and G.1-G.3.*

If for each N $(N = 1, 2, \ldots)$ $(\hat{h}_N, \hat{F}_N) \in (W \times \Gamma)$ *maximizes the likelihood of* N *independent observations* $x^{(N)}$ *over* $(W \times \Gamma)$ *then*

$$Pr\ (\ \lim_{N\to\infty} d_w(\hat{h}_N,\ h*) = 0 \quad \text{and} \quad \lim_{N\to\infty} d_\Gamma(\hat{F}_N,\ F*) = 0\ ) = 1\ .$$

THEOREM 4: *Suppose that the binary threshold model satisfies Assumptions W.1-W.3, W.5′, W.6, Γ.1-Γ.3, and G.1-G.3.*

If for each N $(N = 1, 2, \ldots)$ $(\hat{h}_N, \hat{F}_N) \in (W \times \Gamma)$ *maximizes the likelihood of* N *independent observations* $x^{(N)}$ *over* $(W \times \Gamma)$ *then*

$$Pr\ (\ \lim_{N\to\infty} d'_w(\hat{h}_N,\ h*) = 0 \quad \text{and} \quad \lim_{N\to\infty} d_\Gamma(\hat{F}_N,\ F*) = 0\ ) = 1\ .$$

THEOREM 5: *Suppose that the binary choice model satisfies Assumptions W.1′-W.3′, W.4-W.5, W.6′, W.7, Γ.1-Γ.3, and G.1′-G.3′.*

If for each N $(N = 1, 2, \ldots)$ $(\hat{V}_N, \hat{F}_N) \in (W \times \Gamma)$ *maximizes the likelihood of* N *independent observations* $x^{(N)}$ *over* $(W \times \Gamma)$ *then*

$$Pr\ (\ \lim_{N\to\infty} d_w(\hat{V}_N,\ V*) = 0 \quad \text{and} \quad \lim_{N\to\infty} d_\Gamma(\hat{F}_N,\ F*) = 0\ ) = 1\ .$$

THEOREM 6: *Suppose that the binary choice model satisfies Assumptions W.1′-W.3′, W.4, W.5′-W.6′, Γ.1-Γ.3, and G.1′-G.3′.*

If for each N $(N = 1, 2, \ldots)$ $(\hat{V}_N, \hat{F}_N) \in (W \times \Gamma)$ *maximizes the likelihood of* N *independent observations* $x^{(N)}$ *over* $(W \times \Gamma)$ *then*

$$Pr\ \{\ \lim_{N\to\infty} d'_W(\hat{V}_N, V^*) = 0 \ \text{ and } \ \lim_{N\to\infty} d_\Gamma(\hat{F}_N, F^*) = 0\ \} = 1\ .$$

The proofs of Theorems 3 and 4 are given in Appendix B. The proofs of Theorems 5 and 6 are similar to the proofs of Theorems 3 and 4 and are therefore omitted.

An example of a set of functions satisfying the assumptions in Theorem 3 (Assumptions W.1-W.3, and W.5-W.7) is the set W of all functions $h : T_t \to R$ that are (i) monotone and concave, (ii) homogeneous of degree one, and (iii) satisfy $h(r^*) = \alpha$, where r^* and α are common to all functions h in W, $T_t = R^K_{++}$, and g is a continuous density function with support T_t. The estimation of h over this set of functions has been described in Section 5. That this function satisfies W.1-W.3 was shown in the discussion of Example 1 in Section 3. By (i)-(iii) these functions satisfy W.6 and W.7. By Lemma C.2 in Appendix C, this set of functions satisfy W.5. An example of a set of functions W satisfying the assumptions of Theorem 4 is the set of functions that satisfy (i)-(iii) above, and possess a compact domain and uniformly bounded subgradients.

Examples of sets of functions that satisfy the assumptions in Theorems 5 and 6 can be constructed in a similar way. The compactness of the set W in binary choice models is guaranteed if the set of functions $V(1,\cdot)$ and the set of functions $V(2,\cdot)$ is compact. Assumption W.6′ holds if for each $V \in W$, $V(1,\cdot)$ and $V(2,\cdot)$ satisfy W.6′.

Suppose, for example, that W is the set of functions $V : A \times S \times Z \to R$ that satisfy the conditions in Example 4 of Section 4 with the additional conditions on the concavity of the functions w , the uniform boundedness of the functions t , and on the finite number of elements of the domain of t, which were imposed in Section 5. That this set satisfies Assumptions W.1′, W.3′, and W.4 was shown in Section 3. For any V in this set, $V(2,\cdot)$ satisfies W.6′ by the additive separability of $V(2,\cdot)$ into the continuously distributed vector s_k . A possible way of guaranteeing that $V(1,\cdot)$ satisfies W.6′ is by imposing a positive uniform lower bound on the first coordinate of all subgradients of w . Monotonicity can be imposed by a bound of 0 on the subgradients. By Lemma C.3 in Appendix C, this set is compact if the set of all w functions is compact and the set of all t functions is compact. Since the domain of the t functions is a finite number of points and the values of these functions are uniformly bounded, the set of all t functions is compact with respect to d_w and d'_w . Compactness of the set of w functions with respect to d_w will hold if, in addition to being concave, they are monotone. If we require that Z be compact and the subgradients of the w functions be uniformly bounded, the set of w functions will be made compact with respect to d'_w .

The above examples and the statements of Theorems 3-6 show that the maximum likelihood estimator of (h^*, F^*) $[(V^*, F^*)]$ is strongly consistent under very general assumptions on the set $(W \times \Gamma)$ to which (h^*, F^*) $[(V^*, F^*)]$ is assumed to belong.

7. SUMMARY

In this paper, we have considered the nonparametric identification and consistent estimation of the pair of functions (h^*, F^*) and (V^*, F^*), respectively, in the binary threshold-crossing and binary choice models. The two models were described, and it was shown that (h^*, F^*) and (V^*, F^*) can be identified in sets $(W \times \Gamma)$ of pairs of nonparametric functions. This identification result requires that all functions in W attain the same values on a subset of their domain and that they map this subset onto the support of F^*.

We gave several examples of sets of functions W satisfying the conditions required for the identification of (h^*, F^*) $[(V^*, F^*)]$. The functions in these sets W can be assumed to satisfy properties that are typically derived from economic theory.

We introduced particular estimation methods for (h^*, F^*) $[(V^*, F^*)]$. These methods are based upon Cosslett (1983) and Matzkin (1987) and proceed by maximizing the likelihood functions over $(W \times \Gamma)$. A two-step constrained optimization procedure is used to compute these estimators. We showed that this maximum likelihood estimator is strongly consistent.

APPENDIX A

In this appendix, we present the proofs of Theorems 1 and 2. The statements of these theorems were presented in Section 3.

PROOF OF THEOREM 1:

If $(h,F) \in (W \times \Gamma)$ and $(h,F) \neq (h^*, F^*)$ then either $F \neq F^*$ or $F = F^*$ and $h \neq h^*$. We will present the proof for each of these two cases separately.

Case 1: $F \neq F^*$.

Since $F \neq F^*$, there exists some $D \subset R$ possessing positive Lebesgue measure such that $\forall c \in D$, $F(c) \neq F^*(c)$. Since F and F^* are increasing there exists a point of continuity $t \in D$ of both F and F^*, with $F(t) \neq F^*(t)$. Then, there exists $\delta > 0$ such that either

(A.1) $\forall t' \in (t-\delta, t+\delta)$, $F(t') < F^*(t')$, or

(A.2) $\forall t' \in (t-\delta, t+\delta)$, $F(t') > F^*(t')$.

Let $r^* \in \text{int}\, \bar{T}$ be such that $h^*(r^*) = h(r^*) = t$. By Assumption W.3 such a r^* exists. By G.2, the continuity of h and h^*, (W.1), and the fact that $\bar{T} \subset T_t$, there exists $\varepsilon > 0$ such that the probability measure of the set $N(r^*, \varepsilon) \cap T_t$ is positive and $\forall r \in N(r^*, \varepsilon) \cap T_t$,

$$h^*(r) \in (t-\delta, t+\delta) \quad \text{and} \quad h(r) \in (t-\delta, t+\delta),$$

where $N(r^*, \varepsilon)$ denotes the neighborhood with center r^* and radius ε. From (A.1) and (A.2) it then follows that $\forall r \in N(r^*, \varepsilon) \cap T_t$

$$\Pr(y = 1|r;\ h,\ f) = F(h(r)) \neq F^*(h^*(r)) = \Pr(y = 1|r;\ h^*,\ F^*) .$$

By the definition of identification, this proves that (h^*, F^*) is identified from any $(h, F) \in (W \times \Gamma)$ for which $F \neq F^*$.

<u>Case 2</u>: $F = F^*$ and $h \neq h^*$.

Since $h \neq h^*$, there exists $B \subset T_t$ possessing positive probability measure such that for all $r^* \in B$ either

(A.3) $\quad h(r^*) > h^*(r^*)$, or

(A.4) $\quad h(r^*) < h^*(r^*)$.

Since F^* is strictly increasing by Assumption $\Gamma.2$ it follows that

$$\forall r \in B, \quad \Pr(y = 1|r;h,F) = F(h(r)) \neq F^*(h^*(r)) = \Pr(y = 1|r;h^*,F^*).$$

Hence (h^*, F^*) is identified from any (h, F) when $h \neq h^*$ and $F = F^*$.

This completes the proof of Theorem 1.

Q.E.D.

PROOF OF THEOREM 2:

Let $(V, F) \in (W \times \Gamma)$ be such that $(V, F) \neq (V^*, F^*)$. Again, we distinguish between two cases.

<u>Case 1</u>: $F \neq F^*$.

Since $F \neq F^*$, and F and F^* are increasing, there exists a point of continuity, $t \in R$, of both F and F^* and a $\delta > 0$ such that either

(A.5) $\quad \forall t' \in (t-\delta,\ t+\delta)\ ,\quad F(t') < F^*(t')$, or

(A.6) $\quad \forall t' \in (t-\delta,\ t+\delta)\ ,\quad F(t') > F^*(t')$.

Consider the set $\bar{T}$ in Assumption W.3. Let $(s^*, z_2^*) \in \bar{T}$ be such that

$$V^*(2, s^*, z_2^*) = t .$$

The existence of (s^*, z_2^*) is guaranteed by Assumption W.3(ii). By Assumption W.3(i),

$$V(2, s^*, z_2^*) = t .$$

By Assumption W.4 it then follows that

$$V^*(1, s^*, \bar{z}) - V^*(2, s^*, z_2^*) = t \quad \text{and}$$

$$V(1, s^*, \bar{z}) - V^*(2, s^*, z_2^*) = t .$$

The analysis is similar for the cases in which the domain of the functions V in W is $A \times Z$, $S \times Z$, or Z. Let $r^* = (s^*, \bar{z}, z_2^*)$. By Assumptions G.2 and W.1, there exists $\varepsilon > 0$ such that the probability measure of the set $N(r^*, \varepsilon) \cap S \times Z^2$ is positive and for all $r = (s, z_1, z_2) \in N(r^*, \varepsilon) \cap S \times Z^2$

$$\text{(A.7)} \qquad \begin{aligned} &V^*(1, s, z_1) - V^*(2, s, z_2) \in (t-\delta, t+\delta) \quad \text{and} \\ &V(1, s, z_1) - V(2, s, z_2) \in (t-\delta, t+\delta) . \end{aligned}$$

From (A.5), (A.6), and (A.7), it follows that

$\forall r = (s, z_1, z_2) \in N(r^*, \varepsilon) \cap S \times Z^2$

$$\begin{aligned} \Pr(y = 1|r;V,F) &= F(V(1,s,z_1) - V(1,s,z_2)) \\ &\neq F^*(V^*(1,s,z_1) - V^*(1,s,z_2)) \\ &= \Pr(y = 1|r; V^*, F^*) . \end{aligned}$$

This proves that (V^*, F^*) is identified from (V,F) whenever $F^* \neq F$.

Case 2: $F = F^*$ and $V \neq V^*$

Since $V \neq V^*$, there exists $(\bar{S} \times \bar{Z}_1 \times \bar{z}_2) \subset S \times Z^2$ possessing positive probability measure such that either

(A.8) for all $(s, z_2) \in \bar{S} \times \bar{Z}_2$, $V(2, s, z_2) \neq V^*(2, s, z_2)$, or

(A.9) for all $(s, z_2) \in \bar{S} \times \bar{Z}_2$, $V(2, s, z_2) = V^*(2, s, z_2)$ and for some $(s', z_1') \in \bar{S} \times \bar{Z}_1$, $V(1, s', z_1') \neq V^*(1, s', z_1')$.

If (A.8) is true, it follows from Assumption W.4 that

(A.10) $V(1, s, \bar{z}) - V(2, s, z_2) \neq V^*(1, s, \bar{z}) - V^*(2, s, z_2)$.

If (A.9) is true, it follows that for any $z_2' \in \bar{Z}_2$

(A.11) $V(1, s', z_1') - V(2, s', z_2') \neq V^*(1, s', z_1') - V^*(2, s', z_2')$.

By Γ.3 it then follows that, for all $(s, z_1, z_2) \in N(r^*, \varepsilon) \cap S \cap Z^2$

$$\begin{aligned} Pr(y = 1|r;V,F) &= F^*(V(1, s, z_1) - V(2, s, z_2)) \\ &\neq F^*(V^*(1, s, z_1) - V^*(2, s, z_2)) \\ &= Pr(y = 1|r; V^*, F^*) . \end{aligned}$$

Hence, (V^*, F^*) is identified from (V, F^*) whenever $V \neq V^*$. This completes the proof of Theorem 2. Q.E.D.

APPENDIX B

In this appendix, we present the proofs of Theorem 3 and 4 of Section 6. The proofs are based upon an adaptation of Wald's (1949) result about the strong consistency of maximum likelihood estimators. To make Wald's result applicable, we need to prove that our models possess the required identification, compactness, continuity, measurability, and integratibility properties. The identification property that is employed in the proof of Theorems 3 and 4 was proved in Theorem 1. The identification property necessary to prove Theorems 5 and 6 was proved in Theorem 2. The compactness, continuity, measurability, and integrability properties employed in the proofs of Theorems 3 and 4 are proved in Lemmas 1-5 below. Similar lemmas with almost identical proofs can be employed to prove Theorems 5 and 6.

Before stating and proving the lemmas, we will introduce some additional notation. First, define the metrics $m : (W \times \Gamma) \times (W \times \Gamma) \rightarrow R_+$ and $m' : (W \times \Gamma) \times (W \times \Gamma) \rightarrow R_+$ by

$$m[(h, F), (h', F')] = d_w(h, h') + d_\Gamma(F, F') \quad \text{and}$$

$$m'[(h, F), (h', F')] = d'_w(h, h') + d_\Gamma(F, F') \ .$$

The metrics d_w, d'_w, and d_Γ were defined in Section 6.

Next, let x denote the vector of observable variables (y,r) and f denote the probability density of x. Then, for any $(h,F) \in (W \times \Gamma)$,

$$f(x;h,f) = g(r)[F(h(r))]^y[1 - F(h(r))]^{(1-y)} \ .$$

The set X will denote the support of f and P* will denote the probability measure induced by $f(\cdot;\ h^*, F^*)$.

Define now the functions f' , f'' , f^* , and f^{**} by

$$f'(x;\ h,\ F,\ \varepsilon) = \sup\{f(x;\ h',\ F') \mid m[(h',\ F'),\ (h,\ F)] < \varepsilon\} ,$$

$$f''(x;\ h,\ F,\ \varepsilon) = \sup\{f(x;\ h',\ F') \mid m'[(h',\ F'),\ (h,\ F)] < \varepsilon\} ,$$

$$f^*(x;\ h,\ F,\ \varepsilon) = \begin{cases} f'(x;\ h,\ F,\ \varepsilon) & \text{if } f'(x;\ h,\ F,\ \varepsilon) \geq 1 \\ 1 & \text{otherwise} , \end{cases}$$

$$f^{**}(x;\ h,\ F,\ \varepsilon) = \begin{cases} f''(x;\ h,\ F,\ \varepsilon) & \text{if } f''(x;\ h,\ F,\ \varepsilon) \geq 1 \\ 1 & \text{otherwise} , \end{cases}$$

for sufficiently small $\varepsilon > 0$.

We next present the Lemmas.

LEMMA 0: *Suppose that* $\{h_k\}_{k=1}^{\infty}$ *is a sequence in* W . *Then,*

(0.1) *if* W *satisfies W.1 and W.7 and* $h_k \to h$ *with respect to* d_w , $h_k(r) \to h(r)$ *for all* $r \in T_t$;

(0.2) *if* $h_k \to h$ *with respect to* d'_w , *then* $h_k(r) \to h(r)$ *for all* $r \in Z$.

PROOF: Statement (0.2) follows from the definition of d'_w . To prove (0.1) we note that by the monotonicity of the functions in W (W.7) and the convergence of $\{h_k\}$ to h w.r.t. d_w , it follows that for all $r \in T_t$ such that r is a point of continuity of h , $h_k(r) \to h(r)$. By W.1 it then follows that this pointwise convergence holds for all $r \in T_t$ .Q.E.D.

LEMMA 1 (compactness): *If W.5 and Γ.1 are satisfied, the set* $(W \times \Gamma)$ *is compact with respect to the metric* m . *If W.5′ and Γ.1 are satisfied, the set* $(W \times \Gamma)$ *is compact with respect to the metric* m' .

PROOF: By Assumptions W.5 and W.5′, W is compact with respect to d_W and d'_W respectively. Hence, it remains to show that Γ is compact with respect to d_Γ . Let $\{F_j\}_{j=1}^{\infty}$ be a sequence in Γ . It suffices to show that $\{F_j\}_{j=1}^{\infty}$ has a convergent subsequence. Let Q be a countable dense subset of R. Since the functions F_j are uniformly bounded we can find a subsequence $\{F_{j_k}\}$ of $\{F_j\}$ and a function $F : Q \to R$ such that for all $q \in Q$, $F_{j_k}(q) \to F(q)$. This can be accomplished by the standard diagonalization process (see, for example, Helly's Selection Theorem in Billingsley (1968, pp. 227).) Define $F : R \to R$ by $F(t) = \inf\{F(q) | q \in Q, t \le q\}$. Then, $F \in \Gamma$ and, for all points $t \in R$ at which F is continuous, $F_{j_k}(t) \to F(t)$. Since F is nonincreasing, it is a.e. continuous, hence $F_{j_k} \to F$ a.e. Let $\varepsilon > 0$ be given; and let K be a large enough compact set in R such that

(L1.1) $$\int_{K^c} \exp(-|t|)dt < \varepsilon/2 .$$

By Lebesgue's dominated convergence theorem (Billingsley (1986), p. 213), for all large enough k ,

(L1.2) $$\int_K |F_{j_k}(t) - F(t)| \exp(-|t|)dt < \varepsilon/2 .$$

From (L1.1), (L1.2), and the definition of d_Γ it follows that $d_\Gamma(F_{j_k}, F)$

$\to 0$, as $k \to \infty$. Hence, $\{F_j\}_{j=1}^{\infty}$ has a convergent subsequence. This completes the proof of this Lemma.

Q.E.D.

LEMMA 2 (continuity): *Suppose that W satisfies W.6 and that convergence in W implies pointwise convergence. If Γ.1 is satisfied and $\{(h_k, F_k)\}_{k=1}^{\infty}$ is a sequence in $(W \times \Gamma)$* such that $(h_k, F_k) \to (h, F) \in (W \times \Gamma)$, then $f(x; h_k, F_k) \to f(x; h, F)$ for all $x \in X$, *except possibly on a subset of X of probability measure zero.*

PROOF: Since convergence of $\{h_k\}$ to h implies that $\forall\ r \in T_t$, $h_k(r) \to h(r)$, it follows (see, for example, Cosslett (1983))[3] that if $h(r)$ is a point of continuity of F , $F_k(h_k(r)) \to F(h(r))$, and hence, by the definition of f , that $f(x; h_k, F_k) \to f(x; h, F)$. Since by Γ.1 F has at most a countable number of discontinuities and by W.6 h possesses an absolutely continuous distribution, the subset of T_t at which convergence is guaranteed has unit probability measure.

Q.E.D.

LEMMA 3 (measurability): *Suppose that W satisfies W.1, either W.5 or W.5′, and that convergence in W implies pointwise convergence. If Γ.1 is satisfied, then for any $(h, F) \in (W \times \Gamma)$ and any $\varepsilon > 0$, the function*

[3] Lemma 1 in Appendix B of Cosslett (1983) states that if $\{F_i\}$ is a sequence in Γ such that $F_i \to F \in \Gamma$, $\{\eta_i\}$ a sequence of real numbers such that $\eta_i \to \eta$, and η is a point of continuity of F , then $F_i(\eta_i) \to F(\eta)$.

$f'(x;\ h,\ F,\ \varepsilon)$ is measurable in x if W.5 is satisfied and $f''(x;\ h,\ F,\ \varepsilon)$ is measurable in x if W.5′ is satisfied.

PROOF: We will present the proof for the case in which Assumption W.4 is satisfied. The proof when Assumption W.4′ is satisfied is almost identical and therefore omitted.

Since by W.5 the set W is compact, there exists a countable dense subset $Q \subset W$. Let Γ^r be the countable set of all step functions in Γ whose (finite number of) jumps are rational and occur at rational points. We will show that, for all $x \in X$,

$$\text{(L3.1)}\quad t = \sup\{f(x;\ h', F') \mid (h', F') \in (W \times \Gamma),\ m[(h, F), (h', F')] < \varepsilon\}$$
$$= \sup\{f(x;\ h_i, F_i) \mid (h_i, F_i) \in (Q \times \Gamma^r),\ m[(h_i, F_i), (h, F)] < \varepsilon\} = u .$$

Clearly, $t \geq u$. Suppose $t > u$; then, for some $(h', F') \in (W \times \Gamma)$, some $\alpha \in R$, and $\forall (h_i, F_i) \in (Q \times \Gamma^r)$, either

$$\text{(L3.2)}\quad F'(h'(r)) > \alpha > F(V(r)) , \text{ or}$$

$$\text{(L3.3)}\quad F'(h'(r)) < \alpha < F(h(r)) .$$

Assume w.l.o.g. that (L3.2) holds. Since Q is dense in W, there exists a sequence $(h^j) \subset Q$ such that $h^j \to h'$ with respect to d_W; hence, $h^j(r) \to h'(r)$. Then (see for example, Cosslett (1983)),[4] there exists a sequence $\{F^k\} \subset \Gamma^r$ and a subsequence $\{h^{j_k}\}$ such that $F^k(h^{j_k}(r)) \to F'(h'(r))$.

---

[4]Lemma 2 in Cosslett (1983) states that if $F \in \Gamma$, $\eta \in R$ and $\{\eta_k\}$ is a sequence converging to η, there exists a sequence $\{F_i\} \subset \Gamma^r$ and a subsequence $\{\eta_{k(i)}\}$ such that $F_i(\eta_{k(i)}) \to F(\eta)$.

This contradicts (L3.2); hence $t = u$.

Since by W.1 h^j is continuous and by Γ.1 F^j is increasing, $F^j(h^j(r))$ is measurable on T ; hence, $f(x; h^j, F^j)$ is measurable on X. Consequently, the supremum is measurable too. By (L3.1) , f' is measurable in x .

Q.E.D.

LEMMA 4: *If Assumption G.3 is satisfied,* $\int_X |\log f(x; h^*, F^*)|\, dP^*(x) < \infty$.

PROOF: By the functional structure of $f(x; h^*, F^*)$

$$\begin{aligned} &\int_X |\log f(x; h^*, F^*)|dP^*(x) \\ &\leq 2\int_{T_t} |\log g(r)|g(r)dx \\ &\quad + \int_{T_t} |\log F^*(h^*(r))|[F^*(h^*(r)]g(r)dr \\ &\quad + \int_{T_t} |\log[1 - F^*(h^*(r))]|[1 - F^*(h^*(r)]g(r)dr \ . \end{aligned}$$

Since the ranges of $F(\cdot)$ and $1-F(\cdot)$ are included in the interval [0,1], and since the function $q(y) = y|\log(y)|$ has a bounded range on that interval, the last two integrals are bounded. The first integral is bounded by Assumption G.3. Hence,

$$\int_X |\log f(x; h^*, F^*)| \quad dP^*(x) < \infty \ .$$

Q.E.D.

LEMMA 5: *Suppose that W satisfies W.1, either W.5 or W.5', and that* *convergence in W implies pointwise convergence. If Assumption G.3 is satisfied, then for sufficiently small $\varepsilon > 0$,*

$$\int_X \log f^*(x;\, h,\, F,\, \varepsilon)dP^*(x) \quad \text{and} \quad \int_Z \log f^{**}(x;\, f,\, F,\, \varepsilon)dP^*(x)$$

are finite.

PROOF: We present the proof of Lemma 5 only for f^* , since the proof for f^{**} follows the same lines. Lemma 4 and the assumptions of Lemma 5 imply that f^* is measurable in x . Let $C = \{x \in X | f'(x;\, h,\, F,\, \varepsilon) \le 1\}$, $D = \{x \in X | f'(x;\, h,\, F,\, \varepsilon) > 1\}$, and $E = \{r \in T_t | g(r) > 1\}$. From the definition of f' and our assumptions on Γ it follows that $\forall x = (s,z) \in D$, $g(r) > 1$.

Hence, since by Lemma 4 $f'(x;\, h,\, F,\, \varepsilon)$ is measurable,

$$\begin{aligned}
&\left|\int_X \log(f^*(x;\, h,\, F,\, \varepsilon))dP^*(x)\right| \\
&= \left|\int_C \log(f^*(x;\, h,\, F,\, \varepsilon))\, dP^*(x) + \int_D \log(f^*(x;\, h,\, F,\varepsilon))dP^*(x)\right| \\
&= \left|\int_D \log(f^*(x;\, h,\, F,\, \varepsilon))dP^*(x)\right| \\
&\le 2\int_E |\log(g(r))|g(r)dr \\
&\quad + \int_E |\log(F(h(r))|F(h(r))g(r)dr \\
&\quad + \int_E |\log[1 - F(h(r))]|[1 - F(h(r))]g(r)dr \ .
\end{aligned}$$

Since the first integral is finite by Assumption G.3 and the second and third integrals are finite because the function $q(y) = y|\log(y)|$ has a bounded range on the interval $[0,1]$,

$$\int_X \log f^*(x;\, h,\, F,\, \varepsilon)dP^*(x) \quad \text{is finite.}$$

Q.E.D.

Theorems 3 and 4 can now be proved from these lemmas by adapting, as in Matzkin (1987), Wald's (1949) result to apply to the case when the abstract parameter spaces are compact.

PROOF OF THEOREMS 3 AND 4: First note that the assumptions of either Theorem 3 or Theorem 4 insure that Theorem 1 and Lemma 0 hold. It then follows that Lemmas 1-5 and Theorem 1 also hold. The conclusions of Theorems 3 and 4 now follow by the same argument given in Matzkin (1987), after substituting the parameter θ in Matzkin (1987) by F . For completeness, we repeat those arguments.

By Lemmas 4 and 5, Theorem 1, and Lemma 1 in Wald (1949), for any (h, F) in $(W \times \Gamma)$ such that $(h,F) \neq (h^*, F^*)$,

(T1.1) $$E \log f(X; h, F) < E \log f(X; h^*, F^*) .$$

By Lemmas 2 and 5, and by Lemma 2 in Wald (1949), for any (h, F) in $(W \times \Gamma)$,

(T2.2) $$\lim_{\varepsilon \to 0} E \log f'(X; h, F, \varepsilon) = E \log f(X; h, F) ,$$

where the expectation is taken with respect to P* .

From (T1.1) and (T1.2) it follows that, for any (h, F) in $(W \times \Gamma)$ such that $(h, F) \neq (h^*, F^*)$, there exists $\varepsilon(h, F) > 0$ such that

(T1.3) $$E \log f'(X; h, F, \varepsilon(h,F)) < E \log f(X; h^*, F^*) .$$

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

$$\text{(T1.4)} \quad \text{Prob}\left\{ \lim_{N\to\infty} \frac{\sup_{(V,F)\in Y} \Pi_{i=1}^{N} f(x^i;\, h,\, F)}{\Pi_{i=1}^{N} f(x^i;\, h^*,\, F^*)} = 0 \right\} = 1 .$$

It is clear that

$$Y \subset \cup_{(h,F)\in Y} S(h,\, F,\, \varepsilon(h,F)) = \cup\{S(h,\, F,\, \varepsilon(h,F)) \,|\, (h,F) \in Y\} ,$$

where $S(h, F, \varepsilon(h, F))$ denotes the sphere in $(W \times \Gamma)$ with center (h, F) and radius $\varepsilon(h, F)$.

Since Y is a closed subset of $(W \times \Gamma)$ and since by Lemma 1 $(W \times \Gamma)$ is compact, Y is a compact set. Hence, there exists a finite sequence $\{(h_1, F_1), (h_2, F_2), \ldots, (h_H, F_H)\}$ in Y, and numbers $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_H$, such that $\varepsilon_k = \varepsilon(h_k, F_k)$ and

$$\text{(T1.5)} \qquad Y \subseteq \cup_{k=1}^{H} S(h_k,\, F_k,\, \varepsilon_k) .$$

From (T1.5) and the definition of f' (see the statement of Lemma 3), it follows that, for all N and all $x^1, \ldots, x^N$,

$$\begin{aligned} &\sup_{(h,F)\in Y} \Pi_{i=1}^{N} f(x^i;\, h,\, F) \\ &\le \Sigma_{k=1}^{H} \sup_{(h,F)\in S(h_k,F_k,\varepsilon_k)} \Pi_{i=1}^{N} f(x^i;\, h,\, F) \\ &\le \Sigma_{k=1}^{H} \Pi_{i=1}^{N} f'(x^i;\, h_k,\, F_k,\, \varepsilon_k) . \end{aligned}$$

Hence,

$$\text{(T1.6)} \quad \frac{\sup_{(h,F) \in Y} \Pi_{i=1}^{N} f(x^i; h, F)}{\Pi_{i=1}^{N} f(x^i; h^*, F^*)}$$

$$\leq \frac{\Sigma_{k=1}^{H} \Pi_{i=1}^{N} f'(x^i; h_k, F_k, \varepsilon_k)}{\Pi_{i=1}^{N} f(x^i; h^*, F^*)}$$

By Kolmogorov's Strong Law of Large Numbers and (T1.3) it follows that, for each $k = 1, \ldots, H$

$$\text{(T1.7)} \quad \text{Prob } \{\lim_{N\to\infty} \sum_{i=1}^{N} [\log f'(x^i; h_k, F_k) - \log f(x^i; h^*, F^*)] = -\infty\} = 1.$$

Hence, for $k = 1, \ldots, H$,

$$\text{(T1.8)} \quad \text{Prob } \left\{ \lim_{N\to\infty} \frac{\Pi_{i=1}^{N} f'(x^i; h_k, F_k)}{\Pi_{i=1}^{N} f(x^i; h^*, F^*)} = 0 \right\} = 1 .$$

By (T1.8) and (T1.6), (T1.4) is proved.

Theorem 2 in Wald (1949) and (T1.4) imply now that

$$\text{Prob } \{\lim_{N\to\infty} r[(\hat{h}_N, \hat{F}_N), (h^*, F^*)] = 0\} = 1 .$$

Q.E.D.

APPENDIX C

In this appendix, we state and prove Lemmas C.1-C.3 and describe Asher's et al. algorithm. Lemma C.1 was employed in Section 5 to describe estimation methods for concave functions. The statements of Lemmas C.2 and C.3 were employed in the examples following the statements of Theorems 3-6 in Section 6. Lemma C.2 establishes the compactness, with respect to the metric d_w, of sets W of functions h that are monotone and concave, homogenous of degree one, and that attain a common value at one point of their domain. Lemma C.3 establishes the compactness, with respect to d_w and d'_w, of sets of functions V that are additively separable into functions that belong to compact sets. Asher's et al. algorithm was referred to in Section 5 as a means of finding the solution to the maximization of a likelihood function over a set of distribution functions.

LEMMA C.1: *Suppose that $z^0, \ldots, r^{N+1}$ are N+2 points in the domain $T \subset R^K$ of real valued functions h. Then, the set of all vectors $(h^1, \ldots, h^N)$, for which there exists a concave, monotone, and homogeneous of degree one function $h : T \to R$ with subgradients bounded by $B \in R^K_+$, is the set of all vectors $(h^0, \ldots, h^{N+1})$ satisfying (5.3), (5.4), and (5.5) in Section 5.*

PROOF: The statement of this Lemma is a modification of Afriat's (1973) results about revealed preference, and it has been shown in Matzkin (1987); we therefore repeat here only the main steps.

If h is concave, then for some subgradients $T^0, \ldots, T^{N+1}$

(C.1.1) $h^i = h^r + T^r(r^i - r^r)$, $i, r = 0, 1, \ldots, N, N+1$,

and if h is homogeneous of degree one,

(C.1.2) $h^r = T^r r^r$, $r = 0, 1, \ldots, N, N+1$.

Substituting (C.1.2) into (C.1.1) we obtain (5.3). By (C.1.2), (5.4) is satisfied. The monotonicity of h and the boundedness of its subgradients imply (5.5). On the other hand, if $h^0, \ldots, h^{N+1}$ and $T^0, \ldots, T^{N+1}$ satisfy (5.3), (5.4) and (5.5), the function $h : Z \to R$ defined by $h(r) = \min\{T^i r \mid i = 0, \ldots, N+1\}$ is a concave and homogeneous of degree one function h such that $h(r^j) = h^j$ $(j = 0, \ldots, N+1)$.

Q.E.D.

LEMMA C.2: *Let $r^* \in T_t$ and $\alpha \in R$ be given. Assume that $T_t \subset R^K_{++}$, and let W be the set of all functions $h : T_t \to R$ that are monotone, concave, homogeneous of degree one, and satisfy $h(r^*) = \alpha$. Then W is compact with respect to d_w .*

PROOF: Let $\{h_k\}_{k=1}^{\infty}$ be a sequence in W. We need to show that $\{h_k\}_{k=1}^{\infty}$ has a convergent subsequence.

Define the function $b: T_t \to R$ by $b(r) = \inf \{ T\alpha \mid r \leq T r^* \}$. Since for any $h \in W$ and any $r \in T_t$,
$0 \leq h(r) \leq h(\gamma r^*) = \gamma\alpha$, where $\gamma = \inf \{ T \mid r \leq T r^* \}$, it follows that for any $h \in W$ and any $r \in T_t$, $h(r) \leq b(r)$. Hence, the functions in W are pointwise bounded.

Let Q denote a countable dense subset in T_t. By the diagonalization process, there exists a subsequence $\{h_{k_v}\}$ of $\{h_k\}$ and a function

$h : Q \to R$ such that for all q in Q, $h_{k_v}(q) \to h(q)$. Extend h to T_t by defining for all $r \in T$, $h(r) = \inf \{ h(q) \mid q \in Q , r \leq q \}$. It is then easy to show that $h \in W$. Moreover, for all $r \in T_t$, $h_{k_v}(r) \to h(r)$, since h is continuous and the h_{k_v} functions are monotone.

Since the functions in W are bounded by zero and the function b,

$$(C.2.1) \quad \int |h_{k_v}(r)-h(r)| \exp(-\|r\|) \, dG(r) \leq \int |b(r)| \exp(-\|r\|) \, dG(r).$$

Let $\varepsilon > 0$ be given, and let K be a large enough compact set in T_t such that

$$(C.2.2) \quad \int_{K^c} |b(r)| \exp(-\|r\|) \, dG(r) < \varepsilon/2.$$

Let $\zeta \in R_+$ be such that $\forall\, r \in K$, $r \leq \zeta\, r^*$. Then, since the functions h_{k_v} and h are monotone, they are uniformly bounded on K by $\zeta\,\beta$. By Lebesgue's dominated convergence theorem (Billingsley (1986), pp.213), it then follows that for all large enough v,

$$(C.2.3) \quad \int_K |h_{k_v}(r)-h(r)| \exp(-\|r\|) \, dG(r) < \varepsilon/2.$$

From (C.2.1)-(C.2.3) it follows that $d_w(h_{k_v}, h) \to 0$ as $v \to \infty$, by the definition of d'_w . Hence, $\{h_k\}$ has a convergent subsequence. This completes the proof of Lemma C.2.

Q.E.D.

LEMMA C.3: *Let S and Z be subsets of R^J and R^K respectively; and let W , T and U be sets of functions $V : S \times Z \to R$, $t : S \to R$, and $w : Z \to R$ respectively. Suppose that for each $V \in W$ there exist $t \in T$ and $w \in U$ such that for all $(s,z) \in S \times Z$*

$$V(s,z) = t(s) + w(z) \ .$$

Then, if T *and* U *are compact with respect to the metric* d_w $[d'_w]$, W *is compact with respect to the metric* d_w $[d'_w]$.

PROOF: Let $\{V_n\}_{n=1}^{\infty}$ be a sequence in W. We need to show that $\{V_n\}_{n=1}^{\infty}$ has a convergent subsequence. Since $V_n \in W$, there exists $t_n \in T$ and $w_n \in U$ such that for all $(s,z) \in S \times Z$, $V_n(s,z) = t_n(s) + w_n(z)$ for n = 1, 2, ... Since T is compact with respect to d_w $[d'_w]$, there exists a subsequence $\{t_{n_i}\}$ of $\{t_n\}$ and a function $t \in T$ such that $t_{n_i} \to t$ with respect to d_w $[d'_w]$. Since U is compact with respect to d_w $[d'_w]$, there exists a subsequence $\{w_{n_k}\}$ of $\{w_{n_i}\}$ and a function $w \in U$ such that $w_{n_k} \to w$ with respect to d_w $[d'_w]$. Since $t_{n_i} \to t$, $t_{n_k} \to t$. Let $V: S \times Z$ be defined by $V(s,z) = t(s) + w(z)$ for all $(s,z) \in S \times Z$. We will show that $V_{n_k} \to V$, which will complete the proof of Lemma C.3.

Suppose first that $t_{n_k} \to t$ and $w_{n_k} \to w$ with respect to d_w. Since

$$\begin{aligned} d_w(V_{n_k},V) &= \int |t_{n_k}(s) + w_{n_k}(z) - t(s) - w(z)|\, e^{-\|(s,z)\|}\, dG(s,z) \\ &\le \int (|t_{n_k}(s) - t(s)| + |w_{n_k}(z) - w(z)|)\, e^{-\|(s,z)\|}\, dG(s,z) \\ &\le \int |t_{n_k}(s) - t(s)|\, e^{-\|s\|} dG(s,z) + \int |w_{n_k}(z) - w(z)|\, e^{-\|z\|}\, dG(s,z), \end{aligned}$$

it then follows that $V_{n_k} \to V$ with respect to d_w.

Suppose now that $t_{n_k} \to t$ and $w_{n_k} \to w$ with respect to d'_w. Note that if for some t and some $(s,z) \in S \times Z$

$$|t_{n_k}(s) + w_{n_k}(z) - t(s) - w(z)| > t \quad \text{then, since}$$

$$|t_{n_k}(s) - t(s)| + |w_{n_k}(z) - w(z)| \geq |t_{n_k}(s) + w_{n_k}(z) - t(s) - w(z)|,$$

either $|t_{n_k}(s) - t(s)| > t$ or $|w_{n_k}(z) - w(z)| > t$.

Hence, if for some $\varepsilon > 0$,

$$G(\{(s,z) \in S \times Z| \ |V_{n_k}(s,z) - V(s,z)| > \varepsilon\}) > 0$$

for infinitely many n_k's , it must be that either

$$G(\{s \in S \ | \ |t_{n_k}(s) - t(s)| > \varepsilon\}) > 0 \quad \text{or}$$

$$G(\{z \in Z \ | \ |w_{n_k}(z) - w(s)| > \varepsilon\}) > 0$$

for infinitely many n_k's. Since neither of these possibilities can hold, $V_{n_k} \to V$ with respect to d'_w.

Q.E.D.

Finally, we describe the main steps of the algorithm developed by Asher's et al. (1950) to obtain a solution to the maximization of a likelihood function over probability distribution functions. We have referred to this algorithm in Section 5.

Asher's et al. (1950) Algorithm:

Suppose that in the maximization described in (5.8) of Section 5, the vector $(h^0, \ldots, h^{N+1})$ is given. Then $\hat{F}^1, \ldots, \hat{F}^N$ that solve (5.8) can be found by the following steps, which were introduced by Asher's et al. (1950):

First, rank order the pairs $(h^1, y^1), \ldots, (h^n, y^n)$ in an increasing sequence according to the first coordinate; if for some k , $h^k = h^{k+1}$, order (h^k, y^k) and (h^{k+1}, y^{k+1}) decreasing in the second coordinate.

Second, group the ordered sequence in the minimal number of groups so that the second coordinates of the elements in each group form a decreasing sequence. Third, assign to each F^i the proportion of 1's in the second coordinates of all elements in the group to which (h^i, y^i) belongs. Fourth, check whether the order of the F^i's is the same as the order of the h^i's ; if this is not satisfied, merge the two consecutive groups in which the violation occurs, and return to the third step.

REFERENCES

ASHER, M., H. D. BRUNK, G. M. EWING, W. T. REID, and E. SILVERMAN (1955), "An Empirical Distribution Function for Sampling with Incomplete Information," Annals of Mathematical Statistics, 26, 641-647.

AFRIAT, S. (1972), "The Theory of International Comparison of Real Income and Prices," in J. D. Daly (ed.) International Comparisons of Prices and Output. New York: NBER, 1972.

BILLINGSLEY, P. (1986), Probability and Measure, John Wiley & Sons, Inc.

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

FIACCO A. F. and McCORMICK G. P. (1968) Nonlinear Programming, Sequential Unconstrained Minimization Techniques, New York: Wiley.

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

GALLANT, A. R. (1982), "Unbiased Determination of Production Technologies," Journal of Econometrics, 20.

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

ICHIMURA, H. (1986), "Estimation of Single Index Models," M.I.T.

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.

__________. (1986), "Identification of Binary Response Models," University of Wisconsin Working Paper 8602.

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), "A Nonparametric Maximum Rank Correlation Estimator," Mimeo, Yale University.

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

WALD, A. (1949), "A Note on the Consistency of the Maximum Likelihood Estimator," Annals of Mathematical Statistics, 20: 595-601.