6-1-1989

A Nonparametric Maximum Rank Correlation Estimator

Rosa L. Matzkin

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 918

A NONPARAMETRIC MAXIMUM RANK
CORRELATION ESTIMATOR

BY

ROSA L. MATZKIN

JULY 1989

A NONPARAMETRIC MAXIMUM RANK CORRELATION ESTIMATOR

by

Rosa L. Matzkin*
Cowles Foundation
Yale University
New Haven, CT 06520

May 1988
Revised: April 1989

ABSTRACT

This paper presents a nonparametric and distribution-free estimator for the function, h^*, of observable exogenous variables, x, in the generalized regression model, $y = G(h^*(x), \mu)$. The method does not require a parametric specification for either the function h^* or for the distribution of the random term μ. The function G is only assumed to be monotone increasing. The estimation proceeds by maximizing a rank correlation criterion (Han (1987)) over a set of functions that are monotone increasing, concave, and homogeneous of degree one; the function h^* is assumed to belong to this set of functions. The estimator is shown to be strongly consistent.

---

* This paper was presented at the Conference on Nonparametric and Semiparametric Methods in Econometrics and Statistics, Duke University, May 1988. I am indebted to Charles Manski for useful discussion and to the referees for their comments, suggestions, and corrections. I am in particular grateful to one of the referees for a long and detailed report. I have also benefitted from conversations with Vassilis Hajivassiliou. The support of the National Science Foundation through Grant No. SES-8720596 is gratefully acknowledged.

1. INTRODUCTION

An increasing number of microeconometric models possess limited dependent variables. These include the popular censored, truncated, threshold-crossing, and discrete-choice models. Maddala (1983) describes several applications of such models.

In limited dependent variable models, the observable dependent variable is a transformation, which is not one-to-one, of a latent unobservable dependent variable. For example, the observable dependent variable may be whether an individual accepts a job offer, while the unobservable dependent variable is his willingness to accept. The latent variable is typically assumed to depend on a function h^* of observable exogenous variables x (e.g. salary, outside income) and an unobservable random term μ .

In the past, the estimation of limited dependent variable models has proceeded by specifying parametric structures for the function h^* of exogenous observable variables and for the conditional distribution of the unobservable random term μ given x . The transformation G relating the values of h^* and μ to the values of the observable dependent variables was completely specified. Consequently, these methods were susceptible to potential inconsistencies that could arise from erroneous specifications for G, h^*, or the distribution of μ .

To avoid inconsistencies due to erroneous specification of the distribution of μ , Manski (1975) pioneered the study of semiparametric estimation methods for limited dependent variable models. Manski showed that the parameters of h^* can be consistently estimated in polychotomous choice models without specifying a parametric structure for the distribution of the random terms. The function h^* was assumed to be linear in a finite dimensional parameter and the random terms were only assumed to be i.i.d. within each observation. For binary choice models, even weaker assumptions on the random terms sufficed. Following Manski's work, several other semiparametric distribution-free methods have been developed. These include, among others, Cosslett (1983), Heckman and Singer (1984), Horowitz (1986), Klein and Spady (1986), Manski (1985), and Powell (1984, 1986a, 1986b).

All the above estimators are robust to the misspecification of μ but assume that h^* is known up to a finite-dimensional parameter vector. The specification of h^* , however, is another important source of potential inconsistency. Matzkin (1987) developed a semiparametric estimation method for polychotomous choice models, which did not require h^* to possess a parametric structure. Instead, the function h^* was assumed only to be monotone and concave. The distribution of μ was assumed to be known up to a finite dimensional parameter.

Matzkin (1988a) developed nonparametric distribution free estimators, for single-threshold crossing and binary choice models. These methods do not require a parametric structure for either the function h^* or the distribution of μ. The function h^* , or the function whose difference is h^*, is assumed to possess certain properties such as monotonicity, concavity, and homogeneity of degree one, or some kind of additive separability. The unobservable random term is assumed to be independent of the observable exogenous vector x. These methods provide an estimator of both the function h^* and the distribution of the random term μ. The transformation G, however, is assumed to be perfectly specified.

Recently, some semiparametric distribution-free methods have been introduced that do not assume that the transformation G is known; hence, they avoid the inconsistencies that could arise from an erroneous specification of this transformation. These works include Han (1987), Ichimura (1986), Powell, Stock, and Stoker (1986), and Stoker (1986). Ichimura assumes that h^* is known up to a finite dimensional parameter vector, while the others require h^* to be linear in a finite dimensional parameter.

In this paper, we introduce an estimator for a model in which the transformation G is unknown and neither the distribution of the unobservable random term μ nor the function h^* is specified parametrically. The function G is assumed to be monotone increasing and nonconstant, and the random term μ is assumed to be independent of the exogenous observable variable x. This new estimator is developed by following a suggestion in Matzkin (1988a, Section 3). The estimator is obtained by modifying Han's (1987) semiparametric distribution-free maximum rank correlation estimator and combining it with Matzkin's (1987) semiparametric estimator for monotone and concave functions. The identification of h^* is obtained by requiring it to belong to a set of functions that are monotone increasing, concave, homogenous of degree one, and attain a known value at a known point of their domain. It is appealing to rely on such assumptions, as opposed to requiring restrictive parametric structures, because the assumptions of monotonicity, concavity, and homogeneity of degree one are often encountered in economic models.

In Section 2, we present the model. Then, in Section 3, the new estimator is introduced. In Section 4, we present the assumptions under which the strong consistency of the estimator is established, and in Section 5, we demonstrate the strong consistency of the estimator. Section 6 summarizes and concludes the paper.

2. THE MODEL

In this model, the value of an observable real variable y is determined by the values of a function h^* and an unobservable random variable μ through a transformation G according to the relationship

$$(1) \quad y_i = G\,(\,h^*(x_i),\ \mu_i)\ .$$

The variable x is a K-dimensional random vector of observable exogenous variables distributed according to a probability measure P_x over a set X, which is defined by $X = \{\ x \in R^K \mid x_k > 0 \quad k=1,\ldots,K\ \}$. The random variable μ is distributed independently of x, over a set U, with a probability measure P_μ. The function G is monotone increasing in each coordinate: for all $\mu \le \mu'$ and $t \le t'$, $G(t,\ \mu) \le G(t',\ \mu)$ and $G(t,\ \mu) \le G(t,\ \mu')$. The set of values that G attains over the set $h^*(X) \times U$ is assumed to possess at least two distinct elements. We assume that h^* is monotone increasing, concave, homogeneous of degree one, and satisfies $h^*(x^*) = \alpha$ for some $x^* \in X$ and some $\alpha > 0$, $\alpha \in R$.

This model includes, among others, some proportional and additive hazard models, censored regression models, and threshold crossing models. Han (1987) shows explicitly how these and other particular models are special cases of (1). The model in (1) also belongs to the class of single index models, since the distribution of y conditional on x coincides with the distribution of y conditional on $h^*(x)$.

This paper is concerned with the problem of estimating the function h^* from N independent observations $\{ (y_i, x_i) \mid i = 1, \ldots, N\}$, without imposing any additional restrictions on the function G and without imposing any parametric structure either on the distribution of μ or on the function h^* . This extends Han's (1987) semiparametric generalized regression model, where he assumes that h^* is linear in a parameter β^* that belongs to the set $\{ \beta \in R^K \mid \|\beta\| = 1 , |\beta_K| > \delta \}$. In this paper, we substitute for Han's linear restriction the assumption that the function h^* is monotone increasing, concave, and homogeneous of degree one. These assumptions appear commonly in economic theory. The restriction that β^* belongs to the set $\{ \beta \in R^K \mid \|\beta\| = 1 , |\beta_K| > \delta \}$ is substituted by the assumption that $h^*(x^*) = \alpha$ for some known $x^* \in X$ and some known $\alpha > 0$, $\alpha \in R$.

3. ESTIMATION

In this section we introduce a fully non-parametric estimation method for the model described in the previous section. The estimator is obtained by maximizing a rank correlation criterion over a set W of nonparametric functions. The functions in W satisfy properties that are necessary for the identification of the function h^* and for the strong consistency of the estimator. Following Matzkin (1988a, Section 4, Example 1), we define W to be the set of all monotone increasing, concave, and homogeneous of degree one functions $h: X \to R$ that are defined on the set $X = \{ x \in R^K \mid x_k > 0 , k=1, \ldots, K \}$ and that satisfy $h(x^*) = \alpha$.

Han's (1987) semiparametric maximum rank correlation estimator for the model $y_i = G(x_i \beta, \mu)$ is defined to be any $\hat{\beta}_N$ that maximizes

$$S_N(\beta) =$$

$$= \binom{N}{2}^{-1} \Sigma_{\rho} [1[x_i\beta > x_j\beta] \, 1[y_i > y_j] + 1[x_i\beta < x_j\beta] \, 1[y_i < y_j]]$$

over the set $\{ \beta \in R^K \mid \|\beta\| = 1, |\beta_K| \geq \delta \}$ ($\delta > 0$ is known) where $1[\cdot]$ is an indicator function and Σ_{ρ} denotes the summation over the $\binom{N}{2}$ combinations of two distinct elements (i,j) from $\{1, \ldots, N\}$.

Our nonparametric maximum rank correlation estimator for the function h^* in model (1) is defined to be any function $\hat{h}_N \in W$ that maximizes the function

(2) $S_N(h) =$

$$= \binom{N}{2}^{-1} \Sigma_{\rho}[\ 1[h(x_i) > h(x_j)]\ 1[y_i > y_j] + 1[h(x_i) < h(x_j)]\ 1[y_i < y_j]\ ]\ .$$

over the set W. Hence, instead of searching over a set of linear functions, we search over a set of nonparametric monotone increasing, concave, and homogeneous of degree one functions.

To find a solution to the maximization of (2) over the set W, we can follow a two-step procedure analogous to that developed in Matzkin (1988a):

First maximize the function

(3) $\bar{S}_N(h_1, \ldots, h_N) =$

$$= \binom{N}{2}^{-1} \Sigma_{\rho}\ [\ 1[h_i > h_j]\ 1[y_i > y_j] + 1[h_i < h_j]\ 1[y_i < y_j]\ ]$$

over all vectors $(h_1, \ldots, h_N)$ and $(\beta_1, \ldots, \beta_N)$ that satisfy

(4) $h_i \le \beta_j\, x_i$ $\quad i, j = 0, 1, \ldots, N, N+1$

(5) $h_i = \beta_i\, x_i$ $\quad i = 0, 1, \ldots, N, N+1,$

(6) $\beta_i \ge 0$ $\quad i = 0, 1, \ldots, N, N+1,$ and

(7) $h_0 = 0,\ h_{N+1} = \alpha,\ x_{N+1} = x^*,\ x_0 = 0.$

Second, employ the solution $(h_1^*, \ldots, h_N^*)$ and $(\beta_1^*, \ldots, \beta_N^*)$ to obtain a monotone increasing, concave, and homogeneous of degree one function in W. The resulting function will possess values h_i^* and subgradients β_i^* at each x_i. (See Section 5 in Matzkin (1988a) for the justification of this procedure and for more details.) A similar two step procedure was developed in Matzkin (1987) to maximize a likelihood function over various sets of monotone and concave functions.

Other nonparametric maximum rank correlation estimators could be constructed by specifying different W sets. For example, the set of "least-concave" functions studied in Matzkin (1988b) provides another set of nonparametric functions over which h^* can be strongly consistently estimated. Also the additive separable functions described in Matzkin (1988a, Section 4, Example 2) can be employed to define a set of nonparametric functions W from which a strongly consistent estimator can be constructed.

4. ASSUMPTIONS

In this section, we present and discuss the assumptions under which the nonparametric maximum rank correlation estimator can be shown to be strongly consistent. The proof follows in Section 5. The convergence of our estimator to h^* is obtained with respect to the metric $d : W \times W \to R$ defined by

$$\forall\ h,\ h' \in W \qquad d(h,h') = \int_X |h(x) - h'(x)|\ e^{-\|x\|}\ dx\ .$$

The following assumptions will be made:

A1: For all i,j μ_i and μ_j are i.i.d.

A2: For all i μ_i is independent of x_i.

A3: For all i,j x_i and x_j are i.i.d.

A4: P_x is absolutely continuous.

A5: The support of P_x is $\{x \in R^K \mid x_k \geq 0 , k = 1, \ldots, K \}$.

A6: For all x_i, x_j in X such that $h^*(x_i) < h^*(x_j)$, there exists $t^* \in R$ such that $Pr_{\mu|x}(y_i \leq t^*) > Pr_{\mu|x}(y_j \leq t^*)$, where $Pr_{\mu|x}$ denotes the probability with respect to μ_i conditional on x_i.

A7: $G: h^*(X) \times U \to R$ is monotone increasing in each coordinate and not constant.

A8: W is the set of monotone increasing, concave, and homogeneous of degree one functions $h: X \to R$ such that $h(x^*) = \alpha$, where $X = \{ x \in R^K \mid x_k > 0, k=1, \ldots, K \}$, $x^* \in X$, and $\alpha > 0$.

A9: $h^* \in W$.

Assumptions A1 and A2 are employed together with A7 to show that $h^*(x_i) < h^*(x_j)$ implies that for all $t \in R$, $P_{\underline{\mu}|\underline{x}} (y_i \leq t) \leq P_{\underline{\mu}|\underline{x}} (y_j \leq t)$, where $P_{\underline{\mu}|\underline{x}}$ denotes the probability with respect to (μ_i, μ_j) conditional on (x_i, x_j). Assumption A6 only makes the implication slightly stronger[1]. Assumption 6, together with Assumptions A1, A2, and A7, implies that $P_{\underline{\mu}|\underline{x}} (y_i < y_j) > P_{\underline{\mu}|\underline{x}} (y_i > y_j)$ whenever $h^*(x_i) < h^*(x_j)$. This is employed to prove that h^* uniquely maximizes $E\ S_N(\cdot)$ over W.

Assumptions A8 and A9 together with A3, A4, and A5 guarantee that for all

---

[1] I am indebted to a referee for suggesting Assumption A6 as a substitute for an assumption made in a previous version of this paper.

h in W such that $h \neq h^*$ there exists a set $X_i \times X_j$ in $X \times X$ of positive probability such that for all $(x_i, x_j) \in X_i \times X_j$, $(h(x_i) - h(x_j))\ (h^*(x_i) - h^*(x_j)) < 0$. This is also employed to prove that h^* uniquely maximizes $E\ S_N(\cdot)$ over W.

Assumptions A3, A4, and A8 imply that, for any h in W, the set $\{ (x_i, x_j) \mid h(x_i) = h(x_j) \}$ has zero probability. This is necessary to prove that h^* uniquely maximizes $E\ S_N(\cdot)$ over W and that certain auxiliary functions are continuous in h a.s. The continuity of $E\ S_N(\cdot)$ on W follows from the a.s. continuity in h and the measurability of some of these auxiliary functions.

Assumption A8 guarantees that the set W is compact and that convergence with respect to d implies pointwise convergence. The compactness of W is necessary to prove the uniform convergence of $S_N(\cdot)$ to $E\ S_N(\cdot)$ and the measurability of auxiliary functions. The pointwise convergence of the functions is needed to show the continuity in h and the measurability of the auxiliary functions.

5. CONSISTENCY

In this section we demonstrate the strong consistency of the nonparametric maximum rank correlation estimator. The result is stated in the following theorem:

THEOREM 1: $\lim_{N\to\infty} d(\hat{h}_N, h^*) = 0 \quad a.s.$

To prove this theorem, we need to borrow the result of Lemma C.2 in Matzkin (1988a):

LEMMA 1: *Suppose that W is a set of functions on X and that W satisfies Assumption A8. Then, W is compact with respect to the metric d.*

The proof proceeds in four steps. The first step defines auxiliary functions and studies their properties. The second step establishes the identification of h^*. The third step shows that the rank correlation function converges almost surely uniformly over W to its expectation. And finally, the fourth step employs the first three steps to establish the conclusion of the theorem.

STEP 1 (definitions and properties of random variables):

For any $h \in W$ and $\delta > 0$, define

$$r_{ij}(h) = 1[y_i > y_j]1[h(x_i) > h(x_j)] + 1[y_i < y_j]1[h(x_i) < h(x_j)] ,$$

$$r(h) = E[r_{ij}(h)] ,$$

$$\bar{s}_{ij}(h, \delta) = \sup_{h' \in B(h,\delta)} (r_{ij}(h') - r(h')) ,$$

$$\underline{s}_{ij}(h, \delta) = \inf_{h' \in B(h,\delta)} (r_{ij}(h') - r(h')) ,$$

$$\bar{s}(h, \delta) = E\, \bar{s}_{ij}(h, \delta) \quad \text{, and}$$

$\underline{s}(h, \delta) = E\, \underline{s}_{ij}\, (h, \delta)$,

where $B(h,\delta) = \{ h' \in W \mid d(h, h') < \delta \}$.

Since G and h are monotone increasing, y_i is measurable in (x_i, μ_i) and therefore, since h is measurable,

(1.a.1) $r_{ij}(h)$ *is measurable.*

By its definition,

(1.a.2) $r_{ij}(h)$ *is uniformly bounded over* (i,j) *and* h .

We next show that

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

Let $\{h_k\}$ be a sequence in W and h be an element in W such that $d(h_k, h) \to 0$. Since by Assumption A8 the functions in W are continuous and monotone, convergence with respect to d implies pointwise convergence (see, for example, Matzkin (1988a, Lemma 0)). It then follows that if $x, x' \in X$ and $h(x) > h(x')$, for large enough k, $h_k(x) > h_k(x')$. Thus, $r_{ij}(\cdot)$ is continuous at h if (x_i,x_j) is such that $h(x_i) > h(x_j)$ or $h(x_i) < h(x_j)$. From this it follows that

$$\{ (x_i, x_j) \mid r_{ij}(\cdot) \text{ is not continuous at } h \}$$
$$\subset \{ (x_i, x_j) \mid h(x_i) = h(x_j) \}.$$

By the continuity and homogeneity of degree one of any function h in W , the latter set has Lebesgue measure zero. Hence, since by Assumption A4 P_x is absolutely continuous, and by Assumption A3 x_i is independent of x_j , Pr $\{ (x_i, x_j) \mid r_{ij}(\cdot)$ is not continuous at h $\}$

$\leq$ Pr ((x_i, x_j) | $h(x_i) = h(x_j)$) = 0 .

This completes the proof of (1.a.3).

Next, we note that by (1.a.1) and (1.a.2)

(1.b.1) r(h) exists and it is finite;

by (1.a.2),

(1.b.2) r(h) is uniformly bounded over h;

and by (1.a.2), (1.a.3), (1.b.1), and Lebesgue Dominated Convergence Theorem,

(1.b.3) r(h) is continuous in h .

To study the properties of $\bar{s}_{ij}$ and $\underline{s}_{ij}$, we note that by Lemma 1, W is compact with respect to d, and by Lemma 0 in Matzkin (1988a) and Assumption A8, convergence in W with respect to d implies pointwise convergence. Hence, by (1.a.1), (1.a.3), and (1.b.3) it follows that

(1.c.1) for all $\delta > 0$ *and* $h \in W$, $\bar{s}_{ij}(h, \delta)$ *and* $\underline{s}_{ij}(h, \delta)$ *are measurable.*

By the definition of $\bar{s}_{ij}$ and $\underline{s}_{ij}$, (1.a.2), and (1.b.2), it follows that

(1.c.2) for all $\delta > 0$, $\bar{s}_{ij}(h, \delta)$ *and* $\underline{s}_{ij}(h, \delta)$ *are uniformly bounded over all (i,j) and h.*

By (1.a.3) and (1.b.3)

(1.c.3) $\lim_{\delta \to 0} \bar{s}_{ij}(h, \delta) = (r_{ij}(h) - r(h))$ *a.s.* and

$\lim_{\delta \to 0} \underline{s}_{ij}(h, \delta) = (r_{ij}(h) - r(h))$ *a.s.*

By (1.c.1) and (1.c.2),

(1.d.1) for all $\delta > 0$ *and* $h \in W$, $\bar{s}(h, \delta)$ *and* $\underline{s}(h, \delta)$ *are finite,*

and by (1.c.2), (1.c.3), and Lebesgue Dominated Convergence Theorem,

(1.d.2) $\lim_{\delta \to 0} \bar{s}(h, \delta) = 0$ and $\lim_{\delta \to 0} \underline{s}(h, \delta) = 0$.

This concludes Step 1.

STEP 2 (Identification): h^* *uniquely maximizes* $r(h)$ *over* W .

First we show that

(2.a) $\forall\, x_i, x_j \in X$,

$$h^*(x_i) < h^*(x_j) \text{ implies that } P_{\underline{\mu}|\underline{x}}(y_j > y_i) < P_{\underline{\mu}|\underline{x}}(y_j < y_i) .$$

To show this, we note that since $h^*(x_i) < h^*(x_j)$ and by Assumption A7 G is monotone, for all $\mu \in U$,

$$G(h^*(x_i), \mu) \le G(h^*(x_j), \mu)$$

Since by Assumption A2 μ_i is independent of x_i and by Assumption A1 μ_i and μ_j are i.i.d., it then follows that for all real t

$$Pr_{\mu|x}(y_i \le t) \ \ge\ Pr_{\mu|x}(y_j \le t).$$

Moreover, by Assumption A6 , there exists t^* such that

$$Pr_{\mu|x}(y_i \le t^*) \ >\ Pr_{\mu|x}(y_j \le t^*).$$

It then follows that, since by Assumption A1 μ_i and μ_j are independent,

$$P_{\underline{\mu}|\underline{x}}(y_i > y_j) \ <\ P_{\underline{\mu}|\underline{x}}(y_i < y_j) .$$

This concludes the proof of (2.a).

Second, we note that by the definition of r_{ij} and (2.a) it follows that

(2.b) if $h \in W$ *and* $x_i, x_j \in X$ *are such that* $h^*(x_i) - h^*(x_j) \neq 0$ *, then*

(2.b.1) $E_{\underline{\mu}|\underline{x}} [r_{ij}(h^*) - r_{ij}(h)] \geq 0$ *if*

$$(h(x_i) - h(x_j))\ (h^*(x_i) - h^*(x_j)) \geq 0 \text{ , and}$$

(2.b.2) $E_{\underline{\mu}|\underline{x}} [r_{ij}(h^*) - r_{ij}(h)] > 0$ *if*

$$(h(x_i) - h(x_j))\ (h^*(x_i) - h^*(x_j)) < 0,$$

where $E_{\underline{\mu}|\underline{x}}$ denotes the expectation with respect to $P_{\underline{\mu}|\underline{x}}$.

Third, we show that

(2.c) if $h \in W$ *and* $h \neq h^*$*, then there exists a subset* $X_i \times X_j$ *of* $X \times X$ *of positive probability and such that for all* $(x_i, x_j) \in X_i \times X_j$,

$$(h(x_i) - h(x_j))\ (h^*(x_i) - h^*(x_j)) < 0 \ .$$

Since $h \neq h^*$, there exists $x \in X$ such that $h(x) \neq h^*(x)$. Assume without loss of generality that $h(x) < h^*(x)$. Since by Assumptions A8 and A9 $h(x^*) = h^*(x^*)$ and both functions are homogeneous of degree one, it follows from the definition of X that there exists $\gamma \in R$ such that $h(x) < h(\gamma x^*) = \gamma h(x^*) = \gamma h^*(x^*) = h^*(\gamma x^*) < h^*(x)$.

Since by Assumptions A8 and A9 both h and h^* are continuous and by Assumption A8 X is open, there exists a neighborhood X_i of x and a neighborhood X_j of γx^* such that for all $x_i \in X_i$ and $x_j \in X_j$

$h(x_i) < h(x_j)$ and $h^*(x_i) > h^*(x_j)$. Hence,

$(h(x_i) - h(x_j))\ (h^*(x_i) - h^*(x_j)) < 0$.

Since by Assumptions A4 and A5 P_x is absolutely continuous and its support is the closure of X, and by Assumption A3 x_i is independent of x_j, it follows that $\Pr(X_i \times X_j) > 0$.

This concludes the proof of (2.c).

Fourth, we note again that

(2.d) for all $h \in W$, $\Pr\{(x_i, x_j) \mid h(x_i) = h(x_j)\} = 0$.

This has been shown in the proof of (1.a.3) in Step 1.

Fifth, we employ (2.a) - (2.d) to show that

(2.e) h^* *uniquely maximizes* $r(h)$ *over* W .

Suppose that $h \in W$ and $h \neq h^*$.

Let $A = \{ (x_i, x_j) \mid h^*(x_i) = h^*(x_j) \}$,

$B = \{ (x_i, x_j) \mid (h(x_i) - h(x_j))\ (h^*(x_i) - h^*(x_j)) < 0 \}$, and

$C = A^c \setminus B$,

where $\setminus$ denotes set subtraction.

By (2.b),

$$E_{\underline{\mu}|\underline{x}} [r_{ij}(h^*) - r_{ij}(h)] > 0 \qquad \text{if } (x_i, x_j) \in B \text{ and}$$

$$E_{\underline{\mu}|\underline{x}} [r_{ij}(h^*) - r_{ij}(h)] \geq 0 \qquad \text{if } (x_i, x_j) \in C .$$

Moreover, the probability measure of B is positive by (2.c) and the probability measure of A is zero by (2.d). Since $X \times X = A \cup B \cup C$, it follows then by the definition of $r(\cdot)$ that

$$\begin{aligned} &[r(h^*) - r(h)] \\ &= E [r_{ij}(h^*) - r_{ij}(h)] \\ &= E_{\underline{x}}\ E_{\underline{\mu}|\underline{x}} [r_{ij}(h^*) - r_{ij}(h)] \\ &> 0. \end{aligned}$$

This concludes the proof of (2.e) and of Step 2.

STEP 3 (uniform convergence): *$S_N(h)$ converges a.s. uniformly to $r(h)$.*

This proof follows standard arguments (e.g. Han (1987), Andrews (1987)). Let $\varepsilon > 0$ and $\eta > 0$ be given. By (1.d.2) for each $h \in W$ there exists $\delta(h) > 0$ such that

$$(3.a) \qquad |\, \bar{s}(h,\ \delta(h)\ )\,| < \varepsilon/2 \qquad \text{and} \qquad |\, \underline{s}(h,\ \delta(h)\ )\,| < \varepsilon/2\ .$$

Clearly, $W \subset \cup_{h \in W} B(h,\delta(h))$. Then, since by Lemma 1 W is compact, there exist $h_1, \ldots, h_L$ such that $W \subset \cup_{\ell=1}^{L} B(h_\ell,\delta(h_\ell))$. For each ℓ , let $\delta(h_\ell)$ be denoted by δ_ℓ. Then

$$(3.b) \qquad W \subset \cup_{\ell=1}^{L} B(h_\ell,\delta_\ell).$$

Let $\eta' = \eta\ /\ (2L)$. By (1.c.1), (1.d.1), and the Strong Law of Large Numbers for U-statistics (Serfling (1980)), there exists N such that for each ℓ,

$$(3.c.1) \quad \text{Prob}\left\{ \left| \binom{m}{2}^{-1} \sum_{\rho} \bar{s}_{ij}(h_\ell,\delta_\ell) - s(h_\ell,\delta_\ell) \right| > \varepsilon/2 \quad \text{for some } m \geq N \right\} < \eta'$$

and

$$(3.c.2) \quad \text{Prob}\left\{ \left| \binom{m}{2}^{-1} \sum_{\rho} \underline{s}_{ij}(h_\ell,\delta_\ell) - s(h_\ell,\delta_\ell) \right| > \varepsilon/2 \quad \text{for some } m \geq N \right\} < \eta'.$$

Employing the definitions of $S_N(\cdot)$, $\bar{s}_{ij}$, $\underline{s}_{ij}$, $\bar{s}$, and $\underline{s}$, (1.c.1), (3.a), (3.b), and (3.c.1)-(3.c.2), it follows that

$$\text{Prob}\left\{ \sup_{h\in W} \left| S_m(h) - r(h) \right| > \varepsilon \quad \text{for some} \quad m \geq N \right\}$$

$$\leq \quad \text{Prob}\left\{ \sup_{h\in W} \; (S_m(h) - r(h)) > \varepsilon \quad \text{for some} \quad m \geq N \right\}$$

$$+ \quad \text{Prob}\left\{ \inf_{h\in W} \; (S_m(h) - r(h)) < -\varepsilon \quad \text{for some} \quad m \geq N \right\}$$

$$\leq \sum_{\ell=1}^{L} \text{Prob}\left\{ \sup_{h'\in B(h_\ell,\delta_\ell)} (S_m(h') - r(h')) > \varepsilon \quad \text{for some} \quad m \geq N \right\}$$

$$+ \sum_{\ell=1}^{L} \text{Prob}\left\{ \inf_{h'\in B(h_\ell,\delta_\ell)} (S_m(h') - r(h')) < -\varepsilon \quad \text{for some} \quad m \geq N \right\}$$

$$= \sum_{\ell=1}^{L} \text{Prob}\left\{ \sup_{h'\in B(h_\ell,\delta_\ell)} \left[\binom{m}{2}^{-1} \sum_{\rho} r_{ij}(h') - r(h') \right] > \varepsilon \quad \text{for some } m \geq N \right\}$$

$$+ \sum_{\ell=1}^{L} \text{Prob}\left\{ \inf_{h'\in B(h_\ell,\delta_\ell)} \left[\binom{m}{2}^{-1} \sum_{\rho} r_{ij}(h') - r(h') \right] < -\varepsilon \quad \text{for some } m \geq N \right\}$$

$$\leq \sum_{\ell=1}^{L} \text{Prob}\left\{ \binom{m}{2}^{-1} \sum_{\rho} \sup_{h'\in B(h_\ell,\delta_\ell)} (r_{ij}(h') - r(h')) > \varepsilon \quad \text{for some } m \geq N \right\}$$

$$+ \sum_{\ell=1}^{L} \text{Prob}\left\{ \binom{m}{2}^{-1} \sum_{\rho} \inf_{h'\in B(h_\ell,\delta_\ell)} (r_{ij}(h') - r(h')) < -\varepsilon \quad \text{for some } m \geq N \right\}$$

$$= \sum_{\ell=1}^{L} \text{Prob}\left\{ \binom{m}{2}^{-1} \sum_{\rho} \bar{s}_{ij}(h_\ell, \delta_\ell) > \varepsilon \quad \text{for some } m \geq N \right\}$$

$$+ \sum_{\ell=1}^{L} \text{Prob}\left\{ \binom{m}{2}^{-1} \sum_{\rho} \underline{s}_{ij}(h_\ell, \delta_\ell) < -\varepsilon \quad \text{for some } m \geq N \right\}$$

$$\leq \sum_{\ell=1}^{L} \text{Prob}\left\{ \binom{m}{2}^{-1} \sum_{\rho} \bar{s}_{ij}(h_\ell, \delta_\ell) - \bar{s}(h_\ell,\delta_\ell) > \varepsilon/2 \quad \text{for some } m \geq N \right\}$$

$$+\sum_{\ell=1}^{L} \text{Prob}\left\{ \binom{m}{2}^{-1} \sum_{\rho} \underline{s}_{ij}(h_\ell, \delta_\ell) - \underline{s}(h_\ell, \delta_\ell) < -\varepsilon/2 \quad \text{for some } m \geq N \right\}$$

$$< (2\ L)\ \eta'$$

$$= \eta \ .$$

Hence,

$$\text{Prob}\left\{ \sup_{h \in W} \left| S_m(h) - r(h) \right| > \varepsilon \quad \text{for some} \quad m \geq N \right\} = 0 \ .$$

This concludes Step 3.

STEP 4 (consistency): $\quad lim_{N\to\infty}\ d(\hat{h}_N, h^*) = 0 \quad a.s.$

This is also proved by standard arguments (e.g. Amemiya (1985)). For each $\eta > 0$ let $\bar{B}(\eta) = W \setminus B(h^*, \eta)$ (where $B(h^*,\eta) = \{ h \in W \mid d(h^*, h) < \eta \}$). Since by Lemma 1 W is compact, $\bar{B}(\eta)$ is compact; therefore, since by (1.b.3) in Step 1 r is continuous in h, there exists $\bar{h}(\eta) \in \bar{B}(\eta)$ such that for all $h \in \bar{B}(\eta)$, $r(\bar{h}(\eta)) \geq r(h)$.

Let $\zeta(\eta) = r(h^*) - r(\bar{h}(\eta))$. Since by Step 2 $r(\cdot)$ is uniquely maximized at h^* and by the definition of $\bar{B}(\eta)$ $h^* \notin \bar{B}(\eta)$, $\zeta(\eta) > 0$. Since by Step 3 $S_N(\cdot)$ converges a.s. to r, uniformly in h, there exists $N(\eta)$ such that for all $N > N(\eta)$ $\sup \mid S_N(h) - r(h) \mid < \zeta(\eta)/2$ a.s. Hence, $\hat{h}_N \in B(h^*,\eta)$ a.s. . Since η was arbitrary, this proves that

$$d(\hat{h}_N, h^*) \to 0 \quad \text{a.s.} \ .$$

This concludes Step 4 and the proof of Theorem 1.

6. CONCLUSION

We have presented a nonparametric distribution-free estimator for the function h^* of exogenous variables x in the generalized regression model $y_i = G(h^*(x_i), \mu_i)$, where G is a monotone-increasing nonconstant function, μ is an unobservable random term distributed independently of x, x is K-dimensional random vector that attains values in the set $X = \{ x \in R^K \mid x > 0 \}$, and h^* is a nonparametric function on X that is monotone increasing, concave, homogenous of degree one, and satisfies $h^*(x^*) = \alpha$.

The estimator is obtained by maximizing a rank correlation function over the set of functions $h: X \to R$ that are monotone increasing, concave, homogenous of degree one, and satisfy $h(x^*) = \alpha$. We have discussed a two-step procedure to calculate the estimator, which has been shown to be strongly consistent under the set of Assumptions A1 - A9 specified in Section 5.

REFERENCES

AMEMIYA, T. (1985), Advanced Econometrics, Harvard University Press, Cambridge,Massachusetts.

ANDREWS, D. W. K. (1987), "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," Econometrica, 55, 1465-1471.

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

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

HOROWITZ, J. L. (1986), "A Distribution-Free Least Squares Estimator for Censored Linear Regression Models," Journal of Econometrics, 32, 59-84.

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 Communication 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.

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

__________. (1988b), "Least-concavity and the Distribution-free estimation of Observationally-equivalent Concave Functions," Cowles Foundation, Yale University.

POWELL, J. L. (1984), "Least Absolute Deviations Estimation for the Censored Regression Model," Journal of Econometrics, 25, 303-325.

__________. (1986a), "Censored Regression Quantile," <u>Journal of Econometrics</u>, 32, 143-155.

__________. (1986b), "Symmetrically Trimmed Least Squares Estimation for Tobit Models," <u>Econometrica</u>, 54, 1435-1460.

POWELL, J. L., J. H. STOCK, and T. M. STOKER (1986), "Semiparametric Estimation of Weighted Average Derivatives," Alfred P. Sloan School of Management, WP# 1793.

SERFLING, R. J. (1980), <u>Approximation Theorems of Mathematical Statistics</u>, Willey, New York.

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