UNIVERSIDAD DE SAN ANDRES

DEPARTAMENTO DE ECONOMIA

Restrictions of Economic Theory in Nonparametric Methods, por **Rosa L. Matzkin**, Northwestern University.

Nonparametric Discrete Choice Models with Unobserved Heterogeneity, por **Richard Briesch**, New York University; **Pradeep Chintagunta**, University of Chicago y **Rosa Matzkin**, Northwestern University.

Chapter 42

RESTRICTIONS OF ECONOMIC THEORY IN NONPARAMETRIC METHODS*

ROSA L. MATZKIN

Northwestern University

Contents

*The support of the NSF through Grants SES-8900291 and SES-9122294 is gratefully acknowledged. I am grateful to an editor, Daniel McFadden, and two referees, Charles Manski and James Powell, for their comments and suggestions. I also wish to thank Don Andrews, Richard Briesch, James Heckman, Bo Honoré, Vrinda Kadiyali, Ekaterini Kyriazidou, Whitney Newey and participants in seminars at the University of Chicago, the University of Pennsylvania, Seoul University, Yomsei University and the conference on Current Trends in Economics, Cephalonia, Greece, for their comments. This chapter was partially written while the author was visiting MIT and the University of Chicago, whose warm hospitality is gratefully appreciated.

Handbook of Econometrics, Volume IV, Edited by R.F. Engle and D.L. McFadden

Abstract

This chapter describes several nonparametric estimation and testing methods for econometric models. Instead of using parametric assumptions on the functions and distributions in an economic model, the methods use the restrictions that can be derived from the model. Examples of such restrictions are the concavity and monotonicity of functions, equality conditions, and exclusion restrictions.

The chapter shows, first, how economic restrictions can guarantee the identification of nonparametric functions in several structural models. It then describes how shape restrictions can be used to estimate nonparametric functions using popular methods for nonparametric estimation. Finally, the chapter describes how to test nonparametrically the hypothesis that an economic model is correct and the hypothesis that a nonparametric function satisfies some specified shape properties.

1. Introduction

Increasingly, it appears that restrictions implied by economic theory provide extremely useful tools for developing nonparametric estimation and testing methods. Unlike parametric methods, in which the functions and distributions in a model are specified up to a finite dimensional vector, in nonparametric methods the functions and distributions are left parametrically unspecified. The nonparametric functions may be required to satisfy some properties, but these properties do not restrict them to be within a parametric class.

Several econometric models, formerly requiring very restrictive parametric assumptions, can now be estimated with minimal parametric assumptions, by making use of the restrictions that economic theory implies on the functions of those models. Similarly, tests of economic models that have previously been performed using parametric structures, and hence were conditional on the parametric assumptions made, can now be performed using fewer parametric assumptions by using economic restrictions. This chapter describes some of the existing results on the development of nonparametric methods using the restrictions of economic theory.

Studying restrictions on the relationship between economic variables is one of the most important objectives of economic theory. Without this study, one would not be able to determine, for example, whether an increase in income will produce an increase in consumption or whether a proportional increase in prices will produce a similar proportional increase in profits. Examples of economic restrictions that are used in nonparametric methods are the concavity, continuity and monotonicity of functions, equilibrium conditions, and the implications of optimization on solution functions.

The usefulness of the restrictions of economic theory on parametric models is by now well understood. Some restrictions can be used, for example, to decrease the variance of parameter estimators, by requiring that the estimated values satisfy the conditions that economic theory implies on the values of the parameters. Some can be used to derive tests of economic models by testing whether the unrestricted parameter estimates satisfy the conditions implied by the economic restrictions. And some can be used to improve the quality of an extrapolation beyond the support of the data.

In nonparametric models, economic restrictions can be used, as in parametric models, to reduce the variance of estimators, to falsify theories, and to extrapolate beyond the support of the data. But, in addition, some economic restrictions can be used to guarantee the identification of some nonparametric models and the consistency of some nonparametric estimators.

Suppose, for example, that we are interested in estimating the cost function a typical, perfectly competitive firm faces when it undertakes a particular project, such as the development of a new product. Suppose that the only available data are independent observations on the price vector faced by the firm for the inputs required to perform the project, and whether or not the firm decides to undertake the project. Suppose that the revenue of the project for the typical firm is distributed independently of the vector of input prices faced by that firm. The firm knows the revenue it can get from the project, and it undertakes the project if its revenue exceeds its cost. Then, using the convexity, monotonicity and homogeneity of degree one[1] properties, that economic theory implies on the cost function, one can identify and estimate both the cost function of the typical firm and the distribution of revenues, without imposing parametric assumptions on either of these functions (Matzkin (1992)). This result requires, for normalization purposes, that the cost is known at one particular vector of input prices.

Let us see how nonparametric estimators for the cost function and the distribution of the revenue in the model described above can be obtained. Let $(x^1, \ldots, x^N)$ denote the observed vectors of input prices faced by N randomly sampled firms possessing the same cost function. These could be, for example, firms with the same R&D technologies. Let y^i equal 0 if the ith sampled firm undertakes the project and equal 1 otherwise ($i = 1, \ldots, N$). Let us denote by $h^*(x)$ the cost of undertaking the project when x is the vector of input prices and let us denote by ε the revenue associated with the project. Note that $\varepsilon \geqslant 0$. The cumulative distribution function of ε will be denoted by F^*. We assume that F^* is strictly increasing over the non-negative real numbers and the support of the probability distribution of x is $\mathbb{R}^K_+$. (Since we are assuming that ε is independent of x, F^* does not depend on x.) According to the model, the probability that $y^i = 1$ given x is $\Pr(\varepsilon \leqslant h^*(x^i)) = F^*(h^*(x^i))$. The homogeneity of degree one of h^* implies that $h^*(0) = 0$. A necessary normalization is imposed by requiring that $h^*(x^*) = \alpha$, where both x^* and α are known; $\alpha \in \mathbb{R}$.

[1] A function $h: \mathrm{X} \to \mathbb{R}$, where $\mathrm{X} \subset \mathbb{R}^K$ is convex, is *convex* if $\forall x, y \in \mathrm{X}$ and $\forall \lambda \in [0, 1]$, $h(\lambda x + (1 - \lambda)y) \leqslant \lambda h(x) + (1 - \lambda)h(y)$; h is *homogeneous of degree one* if $\forall x \in \mathrm{X}$ and $\forall \lambda \geqslant 0$, $h(\lambda x) = \lambda h(x)$.

Nonparametric estimators for h^* and F^* can be obtained as follows. First, one estimates the values that h^* attains at each of the observed points $x^1, \ldots, x^N$ and one estimates the values that F^* attains at $h^*(x^1), \ldots, h^*(x^N)$. Second, one interpolates between these values to obtain functions $\hat{h}$ and $\hat{F}$ that estimate, respectively, h^* and F^*. The nonparametric functions $\hat{h}$ and $\hat{F}$ satisfy the properties that h^* and F^* are known to possess. In our model, these properties are that $h^*(x) = \alpha$, h^* is convex, homogeneous of degree one and monotone increasing, and F^* is monotone increasing and its values lie in the interval $[0, 1]$.

The estimator for the finite dimensional vector $\{h^*(x^1), \ldots, h^*(x^N); F^*(h^*(x^1)), \ldots, F^*(h^*(x^N))\}$ is obtained by solving the following constrained maximization log-likelihood problem:

$$\underset{\{F^i\},\{h^i\},\{T^i\}}{\text{maximize}} \sum_{i=1}^{N} \{y^i \log(F^i) + (1 - y^i)\log(1 - F^i)\} \tag{1}$$

subject to

$$F^i \leqslant F^j \quad \text{if} \quad h^i \leqslant h^j, \qquad i, j = 1, \ldots, N, \tag{2}$$

$$0 \leqslant F^i \leqslant 1, \qquad i = 1, \ldots, N, \tag{3}$$

$$h^i = T^i \cdot x^i, \qquad i = 0, \ldots, N+1, \tag{4}$$

$$h^j \geqslant T^i \cdot x^j, \qquad i, j = 0, \ldots, N+1, \tag{5}$$

$$T^i \geqslant 0, \qquad i = 0, \ldots, N+1. \tag{6}$$

In this problem, h^i is the value of a cost function h at x^i, T^i is the subgradient[2] of h at x^i, and F^i is the value of a cumulative distribution at h^i ($i = 1, \ldots, N$); $x^0 = 0$, $x^{N+1} = x^*$, $h^0 = 0$, and $h^{N+1} = \alpha$. The constraints (2)–(3) on $F^1, \ldots, F^N$ characterize the behavior that any distribution function must satisfy at any given points $h^1, \ldots, h^N$ in its domain. As we will see in Subsection 3.1, the constraints (4)–(6) on the values $h^0, \ldots, h^{N+1}$ and vectors $T^0, \ldots, T^{N+1}$ characterize the behavior that the values and subgradients of any convex, homogeneous of degree one, and monotone function must satisfy at the points $x^0, \ldots, x^{N+1}$.

Matzkin (1993b) provides an algorithm to find a solution to the constrained optimization problem above. The algorithm is based on a search over randomly drawn points $(\underline{h}, \underline{T}) = (h^1, \ldots, h^N; T^0, \ldots, T^{N+1})$ that satisfy (4)–(6) and over convex combinations of these points. Given any point $(\underline{h}, \underline{T})$ satisfying (4)–(6), the optimal values of $F^1, \ldots, F^N$ and the optimal value of the objective function given $(\underline{h}, \underline{T})$ are calculated using the algorithm developed by Asher et al. (1955). (See also Cosslett (1983).) This algorithm divides the observations in groups, and assigns to each F^i in a group the value equal to the proportion of observations within the group with

[2] If $f: X \to \mathbb{R}^K$ is a convex function on a convex set $X \subset \mathbb{R}^K$ and $x \in X$, any vector $T \in \mathbb{R}^K$, such that $\forall y \in X\ h(y) \geqslant h(x) + T \cdot (y - x)$, is called a subgradient of h at x. If h is differentiable at x, the gradient of h at x is the unique subgradient of h at x.

$y^i = 1$. The groups are obtained by first ordering the observations according to the values of the h^i's. A group ends at observation i in the jth place and a new group starts at observation k in the $(j+1)$th place iff $y^i = 0$ and $y^k = 1$. If the values of the F^i's corresponding to two adjacent groups are not in increasing order, the two groups are merged. This merging process is repeated till the values of the F^i's are in increasing order. To randomly generate points $(\underline{h}, \underline{T})$, several methods can be used, but the most critical one proceeds by drawing $N+2$ homogeneous and monotone linear functions and then letting $(\underline{h}, \underline{T})$ be the vector of values and subgradients of the function that is the maximum of those $N+2$ linear functions. The coefficients of the $N+2$ linear functions are drawn so that one of the functions attains the value α at x^* and the other functions attain a value smaller than α at x^*.

To interpolate between solution $(\hat{h}^1, \ldots, \hat{h}^N; \hat{T}^0, \ldots, \hat{T}^{N+1}; \hat{F}^1, \ldots, \hat{F}^N)$, one can use different interpolation methods. One possible method proceeds by interpolating linearly between $\hat{F}^1, \ldots, \hat{F}^N$ to obtain a function $\hat{F}$ and using the following interpolation for $\hat{h}$:

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

Figure 1 presents some value sets of this nonparametric estimator $\hat{h}$ when $x \in \mathbb{R}^K_+$. For contrast, Figure 2 presents some value sets for a parametric estimator for h^* that is specified to be linear in a parameter β and x.

At this stage, several questions about the nonparametric estimator described above may be in the reader's mind. For example, how do we know whether these estimators are consistent? More fundamentally, how can the functions h^* and F^* be identified when no parametric specification is imposed on them? And, if they are identified, is the estimation method described above the only one that can be used to estimate the nonparametric model? These and several other related questions will be answered for the model described above and for other popular models.

In Section 2 we will see first what it means for a nonparametric function to be identified. We will also see how restrictions of economic theory can be used to identify nonparametric functions in three popular types of models.

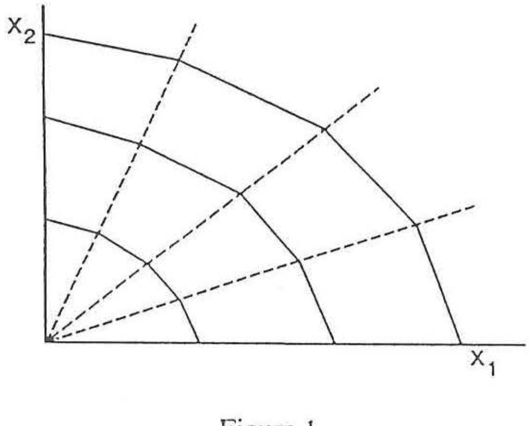

Figure 1

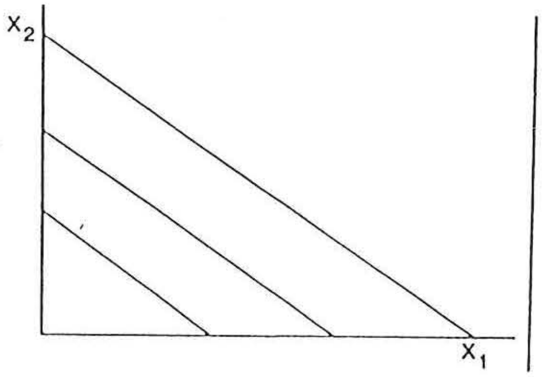

Figure 2

In Section 3, we will consider various methods for estimating nonparametric functions and we will see how properties such as concavity, monotonicity, and homogeneity of degree one can be incorporated into those estimation methods. Besides estimation methods like the one described above, we will also consider seminonparametric methods and weighted average methods.

In Section 4, we will describe some nonparametric tests that use restrictions of economic theory. We will be concerned with both nonstatistical as well as statistical tests. The nonstatistical tests assume that the data is observed without error and the variables in the models are nonrandom. Samuelson's Weak Axiom of Revealed Preference is an example of such a nonparametric test.

Section 5 presents a short summary of the main conclusions of the chapter.

2. Identification of nonparametric models using economic restrictions

2.1. *Definition of nonparametric identification*

Formally, an econometric model is specified by a vector of functionally dependent and independent observable variables, a vector of functionally dependent and independent unobservable variables, a set of known functional relationships among the variables, and a set of restrictions on the unknown functions and distributions. In the example that we have been considering, the observable and unobservable independent variables are, respectively, $x \in \mathbb{R}^K_+$ and $\varepsilon \in \mathbb{R}_+$. A binary variable, y, that takes the value zero if the firm undertakes the project and takes the value 1 otherwise is the observable dependent variable. The profit of the firm if it undertakes the project is the unobservable dependent variable, y^*. The known functional relationships among these variables are that $y^* = \varepsilon - h^*(x)$ and that $y = 0$ when $y^* > 0$ and $y = 1$ otherwise. The restrictions on the functions and distributions are that h^* is continuous, convex, homogeneous of degree one, monotone increasing and attains the value α at x^*; the joint distribution, G, of (x, ε) has as its support the set $\mathbb{R}^{K+1}_+$ and it is such that ε and x are independently distributed.

The restrictions imposed on the unknown functions and distributions in an econometric model define the set of functions and distributions to which these belong. For example, in the econometric model described above, h^* belongs to the set of continuous, convex, homogeneous of degree one, monotone increasing functions that attain the value α at x^*, and G belongs to the set of distributions of (x, ε) that have support $\mathbb{R}_+^{K+1}$ and satisfy the restriction that x and ε are independently distributed.

One of the main objectives of specifying an econometric model is to uncover the "hidden" functions and distributions that drive the behavior of the observable variables in the model. The identification analysis of a model studies what functions, or features of functions, can be recovered from the joint distribution of the observable variables in the model.

Knowing the hidden functions, or some features of the hidden functions, in a model is necessary, for example, to study properties of these functions or to predict the behavior of other variables that are also driven by these functions. In the model considered in the introduction, for example, one can use knowledge about the cost function of a typical firm to infer properties of the production function of the firm or to calculate the cost of the firm under a nonperfectly competitive situation.

Let M denote a set of vectors of functions such that each function and distribution in an econometric model corresponds to a coordinate of the vectors in M. Suppose that the vector, m^*, whose coordinates are the true functions and distribution in the model belongs to M. We say that we can identify within M the functions and distributions in the model, from the joint distribution of the observable variables, if no other vector m in M can generate the same joint distribution of the observable variables. We next define this notion formally.

Let m^* denote the vector of the unknown functions and distributions in an econometric model. Let M denote the set to which m^* is known to belong. For each $m \in M$ let $P(m)$ denote the joint distribution of the observable variables in the model when m^* is substituted by m. Then, the vector of functions m^* is *identified within M* if for any vector $m \in M$ such that $m \neq m^*$, $P(m) \neq P(m^*)$.

One may consider studying the recoverability of some feature, $C(m^*)$, of m^*, such as the sign of some coordinate of m^*, or one may consider the recoverability of some subvector, m_1^*, of m^*, where $m^* = (m_1^*, m_2^*)$. A feature is identified if a different value of the feature generates a different probability distribution of the observable variables. A subvector is identified if, given any possible remaining unknown functions, any subvector that is different can not generate the same joint distribution of the observable variables.

Formally, the feature $C(m^*)$ of m^* is *identified within* the set $\{C(m) | m \in M\}$ if $\forall m \in M$ such that $C(m) \neq C(m^*)$, $P(m) \neq P(m^*)$. The subvector m_1^* is *identified within* M_1, where $M = M_1 \times M_2$, $m_1^* \in M_1$, and $m_2^* \in M_2$, if $\forall m_1 \in M_1$ such that $m_1 \neq m_1^*$, it follows that $\forall m_2, m_2' \in M_2 \quad P(m_1^*, m_2') \neq P(m_1, m_2)$.

When the restrictions of an econometric model specify all functions and distributions up to the value of a finite dimensional vector, the model is said to be parametric. When some of the functions or distributions are left parametrically unspecified, the model is said to be semiparametric. The model is nonparametric if none of the functions and distributions are specified parametrically. For example, in a nonparametric model, a certain distribution may be required to possess zero mean and finite variance, while in a parametric model the same distribution may be required to be a Normal distribution.

Analyzing the identification of a nonparametric econometric model is useful for several reasons. To establish whether a consistent estimator can be developed for a specific nonparametric function in the model, it is essential to determine first whether the nonparametric function can be identified from the population behavior of observable variables. To single out the recoverability properties that are solely due to a particular parametric specification being imposed on a model, one has to analyze first what can be recovered without imposing that parametric specification. To determine what sets of parametric or nonparametric restrictions can be used to identify a model, it is important to analyze the identification of the model first without, or with as few as possible, restrictions.

Imposing restrictions on a model, whether they are parametric or nonparametric, is typically not desirable unless those restrictions are justified. While some amount of unjustified restrictions is typically unavoidable, imposing the restrictions that economic theory implies on some models is not only desirable but also, as we will see, very useful.

Consider again the model of the firm that considers whether to undertake a project. Let us see how the properties of the cost function allow us to identify the cost function of the firm and the distribution of the revenue from the conditional distribution of the binary variable y given the vector of input prices x. To simplify our argument, let us assume that F^* is continuous. Recall that F^* is assumed to be strictly increasing and the support of the probability measure of x is $\mathbb{R}^K_+$. Let $g(x)$ denote $\Pr(y=1|x)$. Then, $g(x)=F^*(h^*(x))$ is a continuous function whose values on $\mathbb{R}^K_+$ can be identified from the joint distribution of (x,y). To see that F^* can be recovered from g, note that since $h^*(x^*)=\alpha$ and h^* is a homogeneous of degree one function, for any $t\in\mathbb{R}_+$, $F^*(t)=F^*((t/\alpha)\,\alpha)=F^*((t/\alpha)\,h^*(x^*))=F^*(h^*((t/\alpha)\,x^*))=g((t/\alpha)\,x^*)$. Next, to see that h^* can be recovered from g and F^*, we note that for any $x\in\mathbb{R}^K_+$, $h^*(x)=(F^*)^{-1}g(x)$. So, we can recover both h^* and F^* from the observable function g. Any other pair (h,F) satisfying the same properties as (h^*,F^*) but with $h\neq h^*$ or $F\neq F^*$ will generate a different continuous function g. So, (h^*,F^*) is identified.

In the next subsections, we will see how economic restrictions can be used to identify other models.

2.2. *Identification of limited dependent variable models*

Limited dependent variable (LDV) models have been extensively used to analyze microeconomic data such as labor force participation, school choice, and purchase of commodities.

A typical LDV model can be described by a pair of functional relationships,

$$y = G(y^*)$$

and

$$y^* = D(h^*(x), \varepsilon),$$

where y is an observable dependent vector, which is a transformation, G, of an unobservable dependent vector, y^*. The vector y^* is a transformation, D, of the value that a function, h^*, attains at a vector of observable variables, x, and the value of an unobservable vector, ε.

In most popular examples, the function D is additively separable into the value of h^* and ε. The model of the firm that we have been considering satisfies this restriction. Popular cases of G are the binary threshold crossing model

$$y = 1 \quad \text{if } y^* \geqslant 0 \text{ and } y = 0 \text{ otherwise,}$$

and the tobit model

$$y = y^* \quad \text{if } y^* \geqslant 0 \text{ and } y = 0 \text{ otherwise.}$$

2.2.1. *Generalized regression models*

Typically, the function h^* is the object of most interest in LDV models, since it aggregates the influence of the vector of observable explanatory variables, x. It is therefore of interest to ask what can be learned about h^* when G and D are unknown and the distribution of ε is also unknown. An answer to this question has been provided by Matzkin (1994) for the case in which y, y^*, $h^*(x)$, and ε are real valued, ε is distributed independently of x, and $G \circ D$ is nondecreasing and nonconstant. Roughly, the result is that h^* is identified up to a strictly increasing transformation. Formally, we can state the following result (see Matzkin (1990b, 1991c, 1994)).

Theorem. Identification of h^* in generalized regression models

Suppose that
- (i) $G \circ D: \mathbb{R}^2 \to \mathbb{R}$ is monotone increasing and nonconstant,
- (ii) $h^*: \mathrm{X} \to \mathbb{R}$, where $\mathrm{X} \subset \mathbb{R}^K$, belongs to a set W of functions $h: \mathrm{X} \to \mathbb{R}$ that are continuous and strictly increasing in the Kth coordinate of x,
- (iii) $\varepsilon \in \mathbb{R}$ is distributed independently of x,
- (iv) the conditional probability of the Kth coordinate of x has a Lebesgue density that is everywhere positive, conditional on the other coordinates of x,
- (v) for any x, x' in X such that $h^*(x) < h^*(x')$ there exists $t \in \mathbb{R}$ such that $\Pr[G \circ D(h^*(x), \varepsilon) \leqslant t] > \Pr[G \circ D(h^*(x'), \varepsilon) \leqslant t]$, where the probability is taken with respect to the probability measure of ε, and
- (vi) the support of the marginal distribution of x includes X.

Then, h^* is identified within W if and only if no two functions in W are strictly increasing transformations of each other.

Assumptions (i) and (iii) guarantee that increasing values of $h^*(x)$ generate non-increasing values of the probability of y given x. Assumption (v) slightly strengthens this, guaranteeing that variations in the value of h^* are translated into variations in the values of the conditional distribution of y given x. Assumption (ii) implies that whenever two functions are not strictly increasing transformations of each other, we can find two neighborhoods at which each function attains different values from the other function. Assumptions (iv) and (vi) guarantee that those neighborhoods have positive probability.

Note the generality of the result. One may be considering a very complicated model determining the way by which an observable vector x influences the value of an observable variable y. If the influence of x can be aggregated by the value of a function h^*, the unobservable random variable ε in the model is distributed independently of x, and both h^* and ε influence y in a nondecreasing way, then one can identify the aggregator function h^* up to a strictly increasing transformation.

The identification of a more general model, where ε is not necessarily independent of x, h^* is a vector of functions, and $G \circ D$ is not necessarily monotone increasing on its domain has not yet been studied.

For the result of the above theorem to have any practicality, one needs to find sets of functions that are such that no two functions are strictly increasing transformations of each other. When the functions are linear in a finite dimensional parameter, say $h(x) = \beta \cdot x$, one can guarantee this by requiring, for example, that $\|\beta\| = 1$ or $\beta_K = 1$, where $\beta = (\beta_1, \ldots, \beta_K)$. When the functions are nonparametric, one can use the restrictions of economic theory.

The set of homogeneous of degree one functions that attain a given value, α, at a given point, x^*, for example, is such that no two functions are strictly increasing transformations of each other. To see this, suppose that h and h' are in this set and for some strictly increasing function f, $h' = f \circ h$; then since $h(\lambda x^*) = h'(\lambda x^*)$ for each $\lambda \geqslant 0$, it follows that $f(t) = f(\alpha(t/\alpha)) = f(h((t/\alpha)x^*)) = h'((t/\alpha)x^*) = t$. So, f is the identity function. It follows that $h' = h$.

Matzkin (1990b, 1993a) shows that the set of least-concave[3] functions that attain common values at two points in their domain is also a set such that no two functions in the set are strictly increasing transformations of each other. The sets of additively separable functions described in Matzkin (1992, 1993a) also satisfy this requirement. Other sets of restrictions that could also be used remain to be studied.

[3]A function $v: X \to \mathbb{R}$, where X is a convex subset of $\mathbb{R}^K$, is *least-concave* if it is concave and if any concave function, v', that can be written as a strictly increasing transformation, f, of v can also be written as a concave transformation, g, of v. For example, $v(x_1, x_2) = (x_1 \cdot x_2)^{1/2}$ is least-concave, but $v(x_1, x_2) = \log(x_1) + \log(x_2)$ is not.

Summarizing, we have shown that restrictions of economic theory can be used to identify the aggregator function h^* in LDV models where the functions D and G are unknown. In the next subsections we will see how much more can be recovered in some particular models where the functions D and G are known.

2.2.2. *Binary threshold crossing models*

A particular case of a generalized regression model where G and D are known is the binary threshold crossing model. This model is widely used not only in economics but in other sciences, such as biology, physics, and medicine, as well. The books by Cox (1970), Finney (1971) and Maddala (1983), among others, describe several empirical applications of these models. The semi- and nonparametric identification and estimation of these models has been studied, among others, by Cosslett (1983), Han (1987), Horowitz (1992), Hotz and Miller (1989), Ichimura (1993), Klein and Spady (1993), Manski (1975, 1985, 1988), Matzkin (1990b, 1990c, 1992), Powell et al. (1989), Stoker (1986) and Thompson (1989).

The following theorem has been shown in Matzkin (1994):

Theorem. Identification of (h^*, F^*) *in a binary choice model*

Suppose that

(i) $y^* = h^*(y) + \varepsilon$; $y = 1$ if $y^* \geqslant 0$, $y = 0$ otherwise.
(ii) $h^*: \mathrm{X} \to \mathbb{R}$, where $\mathrm{X} \subset \mathbb{R}^K$, belongs to a set W of functions $h: \mathrm{X} \to \mathbb{R}$ that are continuous and strictly increasing in the Kth coordinate to x,
(iii) ε is distributed independently of x,
(iv) the conditional probability of the Kth coordinate of x has a Lebesgue density that is everywhere positive, conditional on the other coordinates of x,
(v) F^*, the cumulative distribution function (cdf) of ε, is strictly increasing, and
(vi) the support of the marginal distribution of x is included in X.

Let Γ denote the set of monotone increasing functions on $\mathbb{R}$ with values in the interval $[0, 1]$. Then, (h^*, F^*) is identified within $(\mathrm{W} \times \Gamma)$ if and only if W is a set of functions such that no two functions in W are strictly increasing transformations of each other.

Assumptions (ii)–(iv) and (vi) are the same as in the previous theorem and they play the same role here as they did there. Assumptions (i) and (v) guarantee that assumptions (i) and (v) in the previous theorem are satisfied. They also guarantee that the cdf F^* is identified when h^* is identified.

Note that the set of functions W within which h^* is identified satisfies the same properties as the set in the previous theorem. So, one can use sets of homogeneous of degree one functions, least-concave functions, and additive separable functions to guarantee the identification of h^* and F^* in binary threshold crossing models.

2.2.3. Discrete choice models

Discrete choice models have been extensively used in economics since the pioneering work of McFadden (1974, 1981). The choice among modes of transportation, the choice among occupations, and the choice among appliances have, for example, been studied using these models. See, for example, Maddala (1983), for an extensive list of empirical applications of these models.

In discrete choice models, a typical agent chooses one alternative from a set $\Lambda = \{1, \ldots, J\}$ of alternatives. The agent possesses an observable vector, $s \in S$, of socioeconomic characteristics. Each alternative j in Λ is characterized by a vector of observable attributes $z_j \in Z$, which may be different for each agent. For each alternative $j \in \Lambda$, the agent's preferences for alternative j are represented by the value of a random function U defined by $U(j) = V^*(j, s, z_j) + \varepsilon_j$, where ε_j is an unobservable random term. The agent is assumed to choose the alternative that maximizes his utility; i.e., he is assumed to choose alternative j iff

$$V^*(j, s, z_j) + \varepsilon_j > V^*(k, s, z_k) + \varepsilon_k, \qquad \text{for } k = 1, \ldots, J; k \neq j.$$

(We are assuming that the probability of a tie is zero.)

The identification of these models concerns the unknown function V^* and the distribution of the unobservable random vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J)$. The observable variables are the chosen alternatives, the vector s of socioeconomic characteristics, and the vector $z = (z_1, \ldots, z_J)$ of attributes of the alternatives. The papers by Strauss (1979), Yellott (1977) and those mentioned in the previous subsection concern the nonparametric and semiparametric identification of discrete choice models.

A result in Matzkin (1993a) concerns the identification of V^* when the distribution of the vector of unobservable variables $(\varepsilon_1, \ldots, \varepsilon_J)$ is allowed to depend on the vector of observable variables $(s, z_1, \ldots, z_J)$. Letting $(\varepsilon_1, \ldots, \varepsilon_J)$ depend on (s, z) is important because there is evidence that the estimators for discrete choice models may be very sensitive to heteroskedasticity of ε (Hausman and Wise (1978)). The identification result is obtained using the assumptions that (i) the $V^*(j, \cdot)$ functions are continuous and the same for all j; i.e. $\exists v^*$ such that $\forall j\ V^*(j, s, z_j) = v^*(s, z_j)$, and (ii), conditional on $(s, z_1, \ldots, z_J)$, the ε_j's are i.i.d.[4] Matzkin (1993a) shows that a sufficient condition for $v^*: S \times Z \to \mathbb{R}$ to be identified within a set of continuous functions W is that for any two functions v, v' in W there exists a vector s such that $v(s, \cdot)$ is not a strictly increasing transformation of $v'(s, \cdot)$. So, for example, when the functions $v: S \times Z \to \mathbb{R}$ in W are such that for each s, $v(s, \cdot)$ is homogeneous of degree one, continuous, convex and attains a value α at some given vector z^*, one can identify the function v^*.

A second result in Matzkin (1993a) extends techniques developed by Yellott (1977)

[4] Mański (1975, 1985) used this conditional independence assumption to analyze the identification of semiparametric discrete choice models.

and Strauss (1979). The result is obtained under the assumption that the distribution of ε is independent of the vector (s, z). It is shown that using shape restrictions on the distribution of ε and on the function V^*, one can recover the distribution of the vector $(\varepsilon_2 - \varepsilon_1, \ldots, \varepsilon_J - \varepsilon_1)$ and the $V^*(j, \cdot)$ functions over some subset of their domain. The restrictions on V^* involve knowing its values at some points and requiring that V^* attains low enough values over some sections of its domain. For example, Matzkin (1993a) shows that when V^* is a monotone increasing and concave function whose values are known at some points, V^* can be identified over some subset of its domain.

The nonparametric identification of discrete choice models under other nonparametric assumptions on the distribution of the ε's remains to be studied.

2.3. *Identification of functions generating regression functions*

Several models in economics are specified by the functional relation

$$y = f^*(x) + \varepsilon, \tag{7}$$

where x and ε are, respectively, vectors of observable and unobservable functionally independent variables, and y is the observable vector of dependent variables.

Under some weak assumptions, the function $f^*: X \to \mathbb{R}$ can be recovered from the joint distribution of (x, y) without need of specifying any parametric structure for f^*. To see this, suppose that $E(\varepsilon|x) = 0$ a.s.; then $E(y|x) = f^*(x)$ a.s. Hence, if f^* is continuous and the support of the marginal distribution of x includes the domain of f^*, we can recover f^*. A similar result can be obtained making other assumptions on the conditional distribution of ε, such as $\text{Median}(\varepsilon|x) = 0$ a.s.

In most cases, however, the object of interest is not a conditional mean (or a conditional median) function f^*, but some "deeper" function, such as a utility function generating the distribution of demand for commodities by a consumer, or a production function generating the distribution of profits of a particular firm. In these cases, one could still recover these deeper functions, as long as they influence f^*. This requires using results of economic theory about the properties that f^* needs to satisfy.

For example, suppose that in the model (7) with $E(\varepsilon|x) = 0$, x is a vector (p, I) of prices of K commodities and income of a consumer, and the function f^* denotes for each (p, I) the vector of commodities that maximizes the consumer's utility function U^* over the budget set $\{z \geqslant 0 | p \cdot z \leqslant I\}$; ε denotes a measurement error. Then, imposing theoretical restrictions on f^* we can guarantee that the preferences represented by U^* can be recovered from f^*. Moreover, since f^* can be recovered from the joint distribution of (y, p, I), it follows that U^* can also be recovered from this distribution. Hence, U^* is identified. The required theoretical restrictions on f^* have been developed by Mas-Colell (1977).

Theorem. Recoverability of utility functions from demand functions (Mas-Colell (1977))

Let W denote a set of monotone increasing, continuous, concave and strictly quasi-concave functions such that no two functions in W are strictly increasing transformations of each other. For any $U \in W$, let $f(p, I; U)$ denote the demand function generated by U, where $p \in \mathbb{R}^K_+$ denotes a vector of prices and $I \in \mathbb{R}_{++}$ denotes a consumer's income. Then, for any U, U' in W, such that $U \neq U'$ one has that $f(\cdot,\cdot; U) \neq f(\cdot,\cdot; U')$.

This result states that different utility functions generate different demand functions when the set of all possible values of the vector (p, I) is $\mathbb{R}^{K+1}_+$. The assumption that the utility functions in the set W are concave is the critical assumption guaranteeing that the same demand function can not be generated from two different utility functions in the set W.

Mas-Colell (1978) shows that, under certain regularity conditions, one can construct the preferences represented by U^* by taking the limit, with respect to an appropriate distance function, of a sequence of preferences. The sequence is constructed by letting $\{p^i, I^i\}_{i=1}^{\infty}$ be a sequence that becomes dense in $\mathbb{R}^{K+1}_+$. For each N, a utility function V_N is constructed using Afriat's (1967a) construction:

$$V_N(z) = \min\{V^i + \lambda^i p^i \cdot (z - z^i),\ i = 1, \ldots, N\},$$

where $z^i = f^*(p^i, I^i)$ and the V^i's and λ^i's are any numbers satisfying the inequalities

$$V^i \leqslant V^j + \lambda^j p^j \cdot (z^i - z^j), \qquad i, j = 1, \ldots, N,$$
$$\lambda^i \geqslant 0, \qquad i = 1, \ldots, N.$$

The preference relation represented by U^* is the limit of the sequence of preference relations represented by the functions V_N as N goes to ∞.

Summarizing, we have shown that using Mas-Colell's (1977) result about the recoverability of utility functions from demand functions, we can identify a utility function from the distribution of its demand.

Following a procedure similar to the one described above, one could obtain nonparametric identification results for other models of economic theory. Brown and Matzkin (1991) followed this path to show that the preferences of heterogeneous consumers in a pure exchange economy can be identified from the conditional distribution of equilibrium prices given the endowments of the consumers.

2.4. *Identification of simultaneous equations models*

Restrictions of economic theory can also be used to identify the structural equations of a system of nonparametric simultaneous equations. In particular, when the functions in the system of equations are continuously differentiable, this could be done by determining what type of restrictions guarantee that a given matrix is of full rank. This matrix is presented in Roehrig (1988).

Following Roehrig, let us describe a system of structural equations by

$$r^*(x, y) - u = 0$$

where $x \in \mathbb{R}^K$, $y, u \in \mathbb{R}^G$, and $r^*: \mathbb{R}^K \times \mathbb{R}^G \to \mathbb{R}^G$; y denotes a vector of observable endogeneous variables, x denotes a vector of observable exogenous variables, and u denotes a vector of unobservable exogenous variables. Let ϕ^* denote the joint distribution of (x, u).

Suppose that (i) $\forall(x, y)$ $\partial r^*/\partial y$ is full rank, (ii) there exists a function π such that $y = \pi(x, u)$, and (iii) ϕ^* is such that u is distributed independently of x. Let (r, ϕ) be another pair satisfying these same conditions. Then, under certain assumptions on the support of the probability measures, Roehrig (1988) shows that a necessary and sufficient condition guaranteeing that $P(r^*, \phi^*) = P(r, \phi)$ is that for all $i = 1, \ldots, G$ and all (x, y) the rank of the matrix

$$\begin{pmatrix} \partial r_i/\partial(x, y) \\ \partial r^*/\partial(x, y) \end{pmatrix}$$

is less than $G + 1$. In the above expression, r_i denotes the ith coordinate function of r and $P(r, \phi)$ is the joint distribution of the observable vectors (x, y), when (r^*, ϕ^*) is substituted with (r, ϕ).

Consider, for example, a simple system of a demand and a supply function described by

$$q = d(I, p, w) + \varepsilon_q,$$
$$p = s(w, q, I) + \varepsilon_s,$$

where q denotes quantity, p denotes price, I denotes the income of the consumers and w denotes input price. Then, using the restrictions of economic theory that $\partial d/\partial w = 0$, $\partial s/\partial I = 0$, $\partial d/\partial I \neq 0$ and $\partial s/\partial w \neq 0$, one can show that both the demand function and the supply function are identified up to additive constants.

Kadiyali (1993) provides a more complicated example where Roehrig's (1988) conditions are used to determine when the cost and demand functions of the firms in a duopolistic market are nonparametrically identified. I am not aware of any other work that has used these conditions to identify a nonparametric model.

3. Nonparametric estimation using economic restrictions

Once it has been established that a function can be identified nonparametrically, one can proceed to develop nonparametric estimators for that function. Several methods exist for nonparametrically estimating a given function. In the following subsections we will describe some of these methods. In particular, we will be concerned with the use of these methods to estimate nonparametric functions subject to restrictions of economic theory. We will be concerned only with independent observations.

Imposing restrictions of economic theory on estimator of a function may be necessary to guarantee the identification of the function being estimated, as in the models described in the previous section. They may also be used to reduce the variance of the estimators. Or, they may be imposed to guarantee that the results are meaningful, such as guaranteeing that an estimated demand function is downwards sloping. Moreover, for some nonparametric estimators, imposing shape restrictions is critical for the feasibility of their use. It is to these estimators that we turn next.

3.1. Estimators that depend on the shape of the estimated function

When a function that one wants to estimate satisfies certain shape properties, such as monotonicity and concavity, one can use those properties to estimate the function nonparametrically. The main practical tool for obtaining these estimators is the possibility of using the shape properties of the nonparametric function to characterize the set of values that it can attain at any finite number of points in its domain. The estimation method proceeds by, first, estimating the values (and possibly the gradients or subgradients) of the nonparametric function at a finite number of points of its domain, and second, interpolating among the obtained values. The estimators in the first step are subject to the restrictions implied by the shape properties of the function. The interpolated function in the second step satisfies those same shape properties.

The estimator presented in the introduction was obtained using this method. In that case, the constraints on the vector $(h^1, \ldots, h^N; T^0, \ldots, T^{N+1})$ of values and subgradients of a convex, homogeneous of degree one, and monotone function were

$$h^i = T^i \cdot x^i, \qquad i = 0, \ldots, N+1, \tag{4'}$$

$$h^j \geqslant T^i \cdot x^j, \qquad i, j = 0, \ldots, N+1, \tag{5'}$$

$$T^i \geqslant 0, \qquad i = 0, \ldots, N+1. \tag{6'}$$

The constraints on the vector $(F^1, \ldots, F^N)$ of values of a cdf were

$$F^i \leqslant F^j \qquad \text{if } h^i < h^j,\ i, j = 1, \ldots, N, \tag{2'}$$

$$0 \leqslant F^i \leqslant 1, \qquad i = 1, \ldots, N. \tag{3'}$$

The necessity of the first set of constraints follows by definition. A function $h: \mathrm{X} \to \mathbb{R}$, where X is an open and convex set in $\mathbb{R}^K$, is convex if and only if for all $x \in \mathrm{X}$ there exists $T(x) \in \mathbb{R}^K$ such that for all $y \in \mathrm{X}$, $h(y) \geqslant h(x) + T(x) \cdot (y - x)$. Let h be a convex function and $T(x)$ a subgradient of h at x; h is homogeneous of degree one if and only if $h(x) = T(x)\cdot x$ and h is monotone increasing if and only if $T(x) \geqslant 0$. Letting $x = x^i$, $y = x^j$, $h(x) = h(x^i)$, $h(y) = h^j$ and $T(x) = T^i$ one gets the above constraints. Conversely, to see that if the vector $(h^0,\ldots,h^{N+1}; T^0,\ldots,T^{N+1})$ satisfies the above constraints with $h^0 = 0$ and $h^{N+1} = \alpha$, then its coordinates must correspond to the values and subgradients at $x^0,\ldots,x^{N+1}$ of some convex, monotone and homogeneous of degree one function, we note that the function $h(x) = \max\{T^i\cdot x \mid i = 0,\ldots,N+1\}$ is one such function. (See Matzkin (1992) for a more detailed discussion of these arguments.)

The estimators for (h^*, F^*) obtained by interpolating the results of the optimization in (1)–(6) are consistent. This can be proved by noting that they are maximum likelihood estimators and using results about the consistency of not-necessarily parametric maximum likelihood estimators, such as Wald (1949) and Kiefer and Wolfowitz (1956). To see that $(\hat{h}, \hat{F})$ is a maximum likelihood estimator, let the set of nonparametric estimators for (h^*, F^*) be the set of functions that solve the problem

$$\max_{(h,F)} L_N(h,F) = \sum_{i=1}^{N} \{y^i \log[F(h(x^i))] + (1-y^i)\log[1 - F(h(x^i))]\}$$

$$\text{subject to } (h,F)\in(\mathrm{H}\times\Gamma), \tag{8}$$

where H is the set of convex, monotone increasing, and homogeneous of degree one functions that attain the value α at x^* and Γ is the set of monotone increasing functions on $\mathbb{R}$ whose values lie in the interval $[0, 1]$. Notice that the value of $L_N(h, F)$ depends on h and F only through the values that these functions attain at a finite number of points. As seen above, the behavior of these values is completely characterized by the restrictions (2)–(6) in the problem in the introduction. Hence, the set of solutions of the optimization problem (8) coincides with the set of solutions obtained by interpolating the solutions of the optimization problem described by (1)–(6). So, the estimators we have been considering are maximum likelihood estimators.

We are not aware of any existing results about the asymptotic distribution of these nonparametric maximum likelihood estimators.

The principles that have been exemplified in this subsection can be generalized to estimate other nonparametric models, using possibly other types of extremum estimators, and subject to different sets of restrictions on the estimated functions. The next subsection presents general results that can be used in those cases.

3.1.1. General types of shape restrictions

Generally speaking, one can interpret the theory behind estimators of the sort described in the previous subsection as an immediate extension of the theory behind parametric M-estimators. When a function is estimated parametrically using a maximization procedure, the function is specified up to the value of some finite dimensional parameter vector $\theta \in \mathbb{R}^L$, and an estimator for the parameter is obtained by maximizing a criterion function over a subset of $\mathbb{R}^L$. When the nonparametric shape restricted method is used, the function is specified up to some shape restrictions and an estimator is obtained by maximizing a criterion function over the set of functions satisfying the specified shape restrictions.

The consistency of these nonparametric shape restricted estimators can be proved by extending the usual arguments to apply to subsets of functions instead of subsets of finite dimensional vectors. For example, the following result, which is discussed at length in the chapter by Newey and McFadden in this volume, can typically be used:

Theorem

Let m^* be a function, or a vector of functions, that belongs to a set of functions M. Let L_N: M $\to \mathbb{R}$ denote a criterion function that depends on the data. Let $\hat{m}_N$ be an estimator for m^*, defined by $\hat{m}_N \in \operatorname{argmax}\{L_N(m) | m \in \mathrm{M}\}$. Assume that the following conditions are satisfied:

(i) The function L_N converges a.s. uniformly over M to a nonrandom continuous function L: M $\to \mathbb{R}$.
(ii) The function m^* uniquely maximizes L over the set M.
(iii) The set M is compact with respect to a metric d.

Then, any sequence of estimators $\{\hat{m}_N\}$ converges a.s. to m^* with respect to the metric d. That is, with probability one, $\lim_{N\to\infty} d(\hat{m}_N, m^*) = 0$.

See the Newey and McFadden chapter for a description of the role played by each of the assumptions, as well as a list of alternative assumptions.

The most substantive assumptions are (ii) and (iii). Depending on the definition of L_N, the identification of m^* typically implies that assumption (ii) is satisfied. The satisfaction of assumption (iii) depends on the definitions of the set M and of the metric d, which measures the convergence of the estimator to the true function. Compactness is more difficult to be satisfied by sets of functions than by sets of finite dimensional parameter vectors. One often faces a trade-off between the strength of the convergence result and the strength of the restrictions on M in the sense that the stronger the metric d, the stronger the convergence result, but the more restricted the set M must be. For example, the set of convex, monotone increasing, and homogeneous of degree one functions that attain the value α at x^* and have a common open domain is compact with respect to the L^1 norm. If, in addition, the functions in this set possess uniformly bounded subgradients, then the set is compact with respect to the supremum norm on any compact subset of their joint domain.

Two properties of the estimation method allow one to transform the problem of finding functions that maximize L_N over M into a finite dimensional optimization problem. First, it is necessary that the function L_N depends on any $m \in \mathrm{M}$ only through the values that m attains at a finite number of points. And second, it is necessary that the values that any function $m \in \mathrm{M}$ may attain at those finite number of points can be characterized by a finite set of inequality constraints. When these conditions are satisfied, one can use standard routines to solve the finite dimensional optimization problem that arises when estimating functions using this method. The second requirement is not trivially satisfied. For example, there is no known finite set of necessary and sufficient conditions on the values of a function at a finite number of points guaranteeing that the function is differentiable and α-Lipschitzian[5] ($\alpha > 0$). In the example given in Section 3.1, the concavity of the functions was critical in guaranteeing that we can characterize the behavior of the functions at a finite number of points.

While the results discussed in this section can be applied to a wide variety of models and shape restrictions, some types of models and shape restrictions have received particular attention. We next survey some of the literature concerning estimation subject to monotonicity and concavity restrictions.

3.1.2. *Estimation of monotone functions*

A large body of literature concerns the use of monotone restrictions to estimate nonparametric functions. Most of this work is summarized in an excellent book by Robertson et al. (1988), which updates results surveyed in a previous book by Barlow et al. (1972). (See also, Prakasa Rao (1983).) The book by Robertson et al. describes results about the computation of the estimators, their consistency, rates of convergence, and asymptotic distributions. Subsection 9.2 in that book is of particular interest. In that subsection the authors survey existing results about monotone restricted estimators for the function f^* in the model

$$y = f^*(x) + \varepsilon,$$

where $E(\varepsilon|x) = 0$ a.s. or $\mathrm{Median}(\varepsilon|x) = 0$. Key papers are Brunk (1970), where the consistency and asymptotic distribution of the monotone restricted least squares estimators for f^* is studied when $E(\varepsilon|x) = 0$ and $x \in [0, 1]$; and Hanson et al. (1973), where consistency is proved when $x \in [0, 1] \times [0, 1]$. Earlier, Asher et al. (1955) had proved some weak convergence results. Recently, Wang (1992) derived the rate of convergence of the monotone restricted estimator for f^* when $E(\varepsilon|x) = 0$ a.s. and $x \in [0, 1] \times [0, 1]$. The asymptotic distribution of the least squares estimator for this latter case is not yet known.

Of course, the general methods described in the previous subsection apply in particular to monotone functions. So, one can use those results to determine the consistency of monotone restricted estimators in a variety of models that may or may not fall into the categories of models that are usually studied. (See, for example, Cosslett (1983) and Matzkin (1990a).)

[5] A function $h: \mathrm{X} \to \mathbb{R}$, where $\mathrm{X} \subset \mathbb{R}^K$, is *$\alpha$-Lipschitzian* ($\alpha > 0$) if $\forall x, y \in \mathrm{X}, |h(x) - h(y)| \leqslant \alpha \| x - y \|$.

3.1.3. Estimation of concave functions

Concavity is a shape property that is often satisfied by functions typically found in economic models, such as utility and production functions. Estimation subject to these restrictions has received considerable attention in statistics as well as in economics.

Hildreth (1954) proposed using the monotonicity of the slope of concave functions to derive a least squares estimator for a univariate concave function f^*, in the model

$$y = f^*(x) + \varepsilon.$$

The scalar x was assumed to attain only a finite number of values: $x^1 < x^2 < \cdots < x^N$. The estimators for the values of f^* at $x^1, \ldots, x^N$ were obtained by solving the problem

$$\underset{\{f^1,\ldots,f^N\}}{\text{minimize}} \sum_{i=1}^{N} w_i(\bar{Y}^i - f^i)^2$$

subject to

$$\frac{f^i - f^{i-1}}{x^i - x^{i-1}} \leqslant \frac{f^{i+1} - f^i}{x^{i+1} - x^i}, \qquad i = 2, \ldots, N-1, \tag{9}$$

where w_i is the number of observations with $x = x^i$ and $\bar{Y}^i$ is the average observed value of y when $x = x^i$. Hildreth's paper concerned the computation of this estimator. Hanson and Pledger (1976) proved consistency of this estimator when $E(\varepsilon|x) = 0$ and the support of the probability measure of x is [0, 1]. Nemirovskii et al. (1983) and Mammen (1991b) studied the rate of convergence. The asymptotic distribution of this estimator was recently found by Wang (1992). Let $\hat{f}(x^0)$ denote the value of the concavity restricted least squares estimator at x^0 and let σ^2 denote the variance of ε. Wang showed that, under certain conditions,

$$n^{2/5}\left|\frac{6}{f''(x^0)\sigma^4}\right|^{1/5}(\hat{f}(x^0) - f(x^0))$$

converges to a random variable, Q, that has a symmetric distribution. Using a Monte Carlo method, Wang showed that Q has a bell shaped density, with a standard deviation of approximately 0.8 and a 95% quantile of approximately 1.8. The asymptotic distribution of this estimator when the dimension of x is bigger than 1 is not yet known.

The computation of a concavity restricted estimator for a nonparametric multivariate function can be obtained by estimating not only the values but also the subgradients of the concave function (Matzkin (1986, 1991a), Balls (1987)). The constraints in (9) become

$$f^i \leqslant f^j + T^j \cdot (x^i - x^j), \qquad i, j = 1, \ldots, N,$$

and the minimization is over the values $\{f^i\}$ and the vectors $\{T^i\}$. To add a monotonicity restriction, one includes the constraints

$$T^i \geqslant 0, \qquad i = 1, \ldots, N.$$

To bound the subgradients by a vector B, or to bound the values of the function by the values of a function b, one uses, respectively, the constraints

$$-B \leqslant T^i \leqslant B, \qquad i = 1, \ldots, N,$$

and

$$-b(x^i) \leqslant f^i \leqslant b(x^i), \qquad i = 1, \ldots, N.$$

Algorithms for the resulting constrained optimization problem were developed by Dykstra (1983) and Goldman and Ruud (1992) for the least squares estimator, and Matzkin (1993b) for general types of objective functions. The algorithms by Dykstra and by Goldman and Ruud are extensions of the method proposed by Hildreth (1954). This algorithm proceeds by solving the problem

$$\underset{\lambda \geqslant 0}{\text{minimize}} \; \| y - A'\lambda \|^2,$$

where A is a matrix whose rows are all vectors $\beta \in \mathbb{R}^N$ with $\beta_i = 1$ (some i), $\beta_k \leqslant 0$ (all $k \neq i$), and $\beta' X = 0$. The rows of the $N \times K$ matrix X are the observed points x^i, the first coordinates of which are ones. This is the dual of the problem of finding the vector $\hat{z}$ that minimizes the sum of squared errors subject to concavity constraints

$$\underset{A \cdot z \leqslant 0}{\text{minimize}} \; \| y - z \|^2.$$

The solution to this problem is $\hat{z} = y - A'\hat{\lambda}$, where $\hat{\lambda}$ is the solution to the dual problem. While the dual problem is minimized over more variables, the constraints are much simpler than those of the primal problem. The algorithm minimizes the objective function over one coordinate of λ at a time, repeating the procedure till convergence.

The consistency of the concavity restricted least squares estimator of a multivariate nonparametric concave function can be proved using the consistency result presented in Section 3.1.1. Suppose, for example, that in the model

$$y = f^*(x) + \varepsilon,$$

$x \in X$, where X is an open and convex subset of $\mathbb{R}^K$, $f^*: X \to \mathbb{R}^q$, and the unobserved vector $\varepsilon \in \mathbb{R}^q$ is distributed independently of x with mean 0 and variance Σ. Let $B \in \mathbb{R}^K_+$ and $b: X \to \mathbb{R}^q$. Assume that f^* belongs to the set, H, of concave functions $f: X \to \mathbb{R}^q$ whose subgradients are uniformly bounded by B and their values satisfy that $\forall x \in X$, $|f(x)| \leqslant b(x)$. Then, H is a compact set, in the sup norm, of equicontinuous functions. So, following the same arguments as in, e.g., Epstein and Yatchew (1985) and Gallant (1987), one can show that the function L_N: H $\to \mathbb{R}$ defined by

$$L_N(f) = \frac{1}{N} \sum_{i=1}^{N} (y_i - f(x_i)'\Sigma^{-1}(y_i - f(x_i))$$

converges a.s. uniformly to the continuous function L: H $\to \mathbb{R}$ defined by

$$L(f) = q + \int (f(x) - f^*(x))'(f(x) - f^*(x))\, \mathrm{d}\mu(x),$$

where μ is the probability measure of x. Since the functions in H are continuous, L is uniquely minimized at f^*. Hence, by the theorem of Subsection 3.1.1 it follows that the least squares estimator is a strongly consistent estimator for f^*.

For an LAD (least absolute deviations) nonparametric concavity restricted estimator, Balls (1987) proposed proving consistency by showing that the distance between the concavity restricted estimator and the true function is smaller than the distance between an unrestricted consistent nonparametric splines estimator (see Section 3.2) and the true function. Matzkin (1986) showed consistency of a nonparametric concavity restricted maximum likelihood estimator using a variation of Wald's (1949) theorem, which uses compactness of the set H. No asymptotic distribution results are known for these estimators.

3.2. *Estimation using seminonparametric methods*

Seminonparametric methods proceed by approximating any function of interest with a parametric approximation. The larger the number of observations available to estimate the function, the larger the number of parameters used in the approximating function and the better the approximation. The parametric approximations are chosen so that as the number of observations increases, the sequence of parametric approximations converges to the true function, for appropriate values of the parameters.

A popular example of such a class of parametric approximations is the set of functions defined by the Fourier flexible form (FFF) expansion

$$g_N(x,\theta) = b'x + x'Cx + \sum_{|k|^* \leq T} a_k e^{ik'x}, \qquad x \in \mathbb{R}^K,$$

where $i = \sqrt{-1}$, $b \in \mathbb{R}^K$, C is a $K \times K$ matrix, $a_k = u_k + iv_k$ for some real numbers u_k and v_k, $k = (k_1, \ldots, k_K)$ is a vector with integer coordinates, and $|k|^* = \sum_{i=1}^{K} |k_i|$. (See Gallant (1981).)

To guarantee that the above sum is real valued, it is imposed that $v_0 = 0$, $u_k = u_{-k}$ and $v_k = -v_{-k}$. Moreover, the values of each coordinate of x need to be modified to fall into the $[0, 2\pi]$ interval. The coordinates of the parameter vector θ are the u_k's, the v_k's and the coefficients of the linear and quadratic terms. Important advantages of this expression are that it is linear in the parameters and its partial derivatives are easily calculated. As $K \to \infty$, the FFF and its partial derivatives up to order $m-1$ approximate in an L^p norm any m times differentiable function and its $m-1$ derivatives.

Imposing restrictions on the values of the parameters of the approximation, one can guarantee that the resulting estimator satisfies a desired shape property. Gallant and Golub (1984), for example, impose quasi-convexity in the FFF estimator by calculating the estimator for θ as the solution to a constrained minimization problem

$$\min_{\theta} s_N(\theta) \quad \text{subject to} \quad r(\theta) \geq 0,$$

where $s_N(\cdot)$ is a data dependent function, such as a weighted sum of squared errors, $r(\theta) = \min_x v(x,\theta)$ and $v(x,\theta) = \min_z \{z'D^2 g_N(x,\theta)z \mid z'Dg_N(x,\theta) = 0, z'z = 1\}$. Dg_N and $D^2 g_N$ denote, respectively, the gradient and Hessian of g_N with respect to x. Gallant and Golub (1984) have developed an algorithm to solve this problem.

Gallant (1981, 1982) developed restrictions guaranteeing that the Fourier flexible form approximation satisfies homotheticity, linear homogeneity or separability.

The consistency of seminonparametric estimators can typically be shown by appealing to the following theorem, which is presented and discussed in Gallant (1987) and Gallant and Nychka (1987, Theorem 0).

Theorem

Suppose that m^* belongs to a set of functions M. Let L_N: M $\to \mathbb{R}$ denote a criterion function that depends on the data. Let $\{M_N\}$ denote an infinite sequence of subsets of M such that $\cdots M_N \subset M_{N+1} \subset M_{N+2} \cdots$. Let m_N^F be an estimator for m^*, defined by $m_N^F = \operatorname{argmax}\{L_N(m) \mid m \in M_N\}$. Assume that the following conditions are satisfied.

(i) The function L_N converges a.s. uniformly over M to a nonrandom continuous function L: M $\to \mathbb{R}$.

(ii) The function m^* uniquely maximizes L over the set M.

(iii) The set M is compact with respect to a metric d.
(iv) There exists a sequence of functions $\{g_N\} \subset \mathrm{M}$ such that $g_N \in \mathrm{M}_N$ for all $N = 1, 2, \ldots$ and $d(g_N, m^*) \to 0$.

Then, the sequence of estimators $\{m_N\}$ converges a.s. to m^* with respect to the metric d. That is, with probability one, $\lim_{N\to\infty} d(m_N, m^*) = 0$.

This result is very similar to the theorem in Subsection 3.1.1. Indeed, Assumptions (i)–(iii) play the same role here as they played in that theorem. Assumption (iv) is necessary to substitute for the fact that the maximization of L_N for each N is not over the whole space M but only over a subset, M_N, of M. This asumption is satisfied when the M_N sets become dense in M as $N \to \infty$. (See Gallant (1987) for more discussion about this result.)

Asymptotic normality results for Fourier flexible forms and other seminonparametric estimators have been developed, among others, by Andrews (1991), Eastwood (1991), Eastwood and Gallant (1991) and Gallant and Souza (1991). None of these considers the case where the estimators are restricted to be concave.

The M_N sets are typically defined by using results that allow one to characterize any arbitrary function as the limit of an infinite sum of parametric functions. The Fourier flexible form described above is one example of this. Each set M_N is defined as the set of functions obtained as the sum of the first $T(N)$ terms in the expansion, where $T(N)$ is increasing in N and such that $K(N) \to \infty$ as $N \to \infty$.

Some other types of expansions that have been used to define parametric approximations are Hermite forms (Gallant and Nychka (1987)), power series (Bergstrom (1985)), splines (Wahba (1990)), and Müntz–Szatz type series (Barnett and Yue (1988a, 1988b) and Barnett et al. (1991)).

Splines are smooth functions that are piecewise polynomials. Kimeldorf and Wahba (1971), Utreras (1984, 1985), Villalobos and Wahba (1987) and Wong (1984) studied the imposition of monotonicity and convexity restrictions on splines estimators. Yatchew and Bos (1992) proposed using splines to estimate a consumer demand function subject to the implications of economic theory on demand functions.

Barnett et al. (1991) impose concavity in a Müntz–Szatz type series by requiring that each term in the expansion satisfies concavity. This method for imposing concavity restrictions in series estimators was proposed by McFadden (1985).

3.3. *Estimation using weighted average methods*

A weighted average estimator, $\hat{f}$, for the function f^* in the model

$$y = f^*(x) + \varepsilon,$$

where $E(\varepsilon|x) = 0$, has the form

$$\hat{f}(x) = \sum_{i=1}^{N} w_{N,i}(x) y^i.$$

In this expression, $\{y^i, x^i\}_j^N$ is a set of N observations and the $w_{N,i}$ functions are such that $\sum_{i=1}^{N} w_{N,i}(x) = 1$. The functions $w_{N,i}$ are typically nonnegative and decreasing in the vector $(x - x^1, \ldots, x - x^N)$, so that observations at nearby points are given more weight in the estimation of $f^*(x)$ than further away points. The sequence of functions $\{w_{N,i}\}$ is chosen so as to guarantee the consistency and asymptotic normality of the estimator. The method of kernels (Nadaraja (1964), Watson (1964)) and the method of nearest neighbors (Royall (1966)) are examples of weighted average methods.

When no particular restrictions are imposed on these estimators, their calculation does not require maximization of a function. This is an advantage over the estimators discussed in the previous sections. The asymptotic distribution of the unrestricted estimators has been extensively studied. See the chapter by Härdle and Linton in this volume for a survey of those results.

To use shape restrictions with weighted average estimators, one can first use a weighted average method to smooth the data and then use a shape restricted method, such as the methods described in Section 3.1.1, with the smoothed data. Or, one can interchange these steps. The estimators developed by Friedman and Tibshirani (1984), Wright (1982) and Mukarjee's (1988) are of these types. Mammen (1991a) studied the behavior of two estimators, m_{SI} and m_{IS}. The m_{SI} estimator results from first obtaining $\hat{f}(x)$ as

$$\hat{f}(x) = \sum_{j=1}^{N} w_N(x - x^j) y^j,$$

and second, calculating a monotone estimator as a solution to

$$\inf_{f \in H} \int (\hat{f}(x) - f(x))^2 \, \mathrm{d}x,$$

where H is a set of monotone functions. The m_{IS} estimator is obtained by reversing the steps. He showed that when the $\{\varepsilon^i\}$ are i.i.d. and the $x^i \in [-1, 1]$ and are equally spaced, the estimators for $f^*(0)$ resulting from either of these procedures are both of the order $N^{-2/5}$, for an appropriate $\{w_N(\cdot)\}$ sequence.

Mukarjee and Stern (1994) proposed using a first-smooth second-monotonize procedure to alleviate the computational burden involved with estimating a multivariate monotonicity restricted least squares estimator. The estimator, $\hat{f}$, is defined by

$$\hat{f}(x) = (\hat{f}_U(x) + \hat{f}_L(x))/2,$$

where

$$\hat{f}_U(x) = \min_{x' \geqslant x} \hat{f}(x'), \quad \hat{f}_L(x) = \max_{x' \leqslant x} \hat{f}_K(x')$$

and where $\hat{f}_K$ is a kernel estimator for f. The consistency of this estimator follows from the consistency of the kernel estimator. No asymptotic distribution for it is known.

A kernel estimator was also used in Matzkin (1991d) to obtain a smooth interpolation of a concavity restricted nonparametric maximum likelihood estimator and in Matzkin and Newey (1992) to estimate a homogenous function in a binary threshold crossing model. The Matzkin and Newey estimator possesses a known asymptotic distribution.

4. Nonparametric tests using economic restrictions

The testing of economic hypotheses in parametric models suffers from drawbacks similar to those of the estimation of parametric models; the conclusions depend on the parametric specifications used. Suppose, for example, that one is interested in testing whether some given consumer demand data provide support for the classical model of utility maximization. The parametric approach would proceed by: first, specifying parametric structures for the demand functions; second, using the demand data to estimate the parameters; and then testing whether the estimated demand functions satisfy the integrability conditions. But, if the integrability conditions are not satisfied by the parametrically estimated demand functions it is not clear whether this is evidence against the utility maximization model *or* against the particular parametric structures chosen. In contrast, a nonparametric test of the utility maximization model would use demand functions estimated nonparametrically. In this case, rejection of the integrability conditions provides stronger evidence against the utility maximization model.

4.1. Nonstatistical tests

A large body of literature dating back to the work of Samuelson (1938) and Houthakker (1950) on Revealed Preference has developed nonparametric tests for the hypothesis that data is consistent with a particular choice model, such as the choice made by a consumer or a firm. Most of these tests are nonstatistical. The data is assumed to be observed without error and the models contain no unobservable random terms. (One exception is the Axiom of Revealed Stochastic Rationality (McFadden and Richter (1970, 1990)), where conditions are given characterizing discrete choice probabilities generated by a random utility function.) In the nonstatistical tests, an hypothesis is rejected if at least one in a set of nonparametric restrictions is violated; the hypothesis is accepted otherwise. The nonparametric restrictions used to test the hypotheses are typically expressed in one of two different ways. Either they establish that a solution must exist for a certain finite system of inequalities whose coefficients are determined by the data; or they establish that certain algebraic conditions must be satisfied by the data. For example, the Strong Axiom of Revealed Preference is one of the algebraic conditions that is used in these tests.

To provide an example of such results, we state below Afriat's (1967a) Theorem, which is fundamental in this literature. Afriat's Theorem can be used to test the consistency of demand data with the hypothesis that observed commodity bundles are the maximizers of a common utility function over the budget sets determined by observed prices of the commodities and incomes of a consumer. If the data correspond to different individuals, the conditions of the theorem can be used to test the existence of a utility function that is common to all of them.

Afriat's Theorem (1967a)

Let $\{x^i, p^i, I^i\}_{i=1}^N$ denote a set of N observations on commodity bundles x^i, prices p^i, and incomes I^i such that $\forall i, p^i \cdot x^i = I^i$. Then, the following conditions are equivalent.

(i) There exists a nonsatiated function $V: \mathbb{R}^K \to \mathbb{R}$ such that for all $i = 1, \ldots, N$ and all $y \in \mathbb{R}^K$, $[p^i \cdot y \leqslant I^i] \Rightarrow [V(y) \leqslant V(x^i)]$.

(ii) The data $\{x^i, p^i, I^i\}_{i=1}^N$ satisfy Cyclical Consistency; i.e., for all sequences $\{i, j, k, \ldots, r, l\}$

$$[p^j \cdot x^i \leqslant I^j, p^k \cdot x^j \leqslant I^k, \cdots, p^l \cdot x^r \leqslant I^l] \Rightarrow [I^i \leqslant p^i \cdot x^l].$$

(iii) There exist numbers $\lambda^i > 0$ and V^i $(i = 1, \ldots, N)$ satisfying

$$V^i \leqslant V^j + \lambda^j p^j \cdot (x^i - x^j), \qquad i, j = 1, \ldots, N.$$

(iv) There exists a monotone increasing, concave and continuous function $V: \mathbb{R}^K \to \mathbb{R}$ such that for all $i = 1, \ldots, N$ and all $y \in \mathbb{R}^K$, $[p^i \cdot y \leqslant I^i] \Rightarrow [V(y) \leqslant V(x^i)]$.

This result states that the data could have been generated by the maximization of a common nonsatiated utility function (condition (i)) if and only if that data satisfy the set of algebraic conditions stated in condition (ii). In Figure 3, two observations that do not satisfy Cyclical Consistency are graphed. In these observations, $p^1 \cdot x^2 < I^1 = p^1 \cdot x^1$ and $p^2 \cdot x^1 < I^2 = p^2 \cdot x^2$.

The theorem also states that a condition equivalent to Cyclical Consistency is that one can find numbers $\lambda^i > 0$ and V^i $(i = 1, \ldots, N)$ satisfying the linear inequalities in (iii). For example, no such numbers can be found for the observations in Figure 3; since when $p^1 \cdot (x^2 - x^1) < 0$, $p^2 \cdot (x^1 - x^2) < 0$, and $\lambda^1, \lambda^2 > 0$, the inequalities in (iii) imply that $V^1 - V^2 < 0$ and $V^2 - V^1 < 0$.

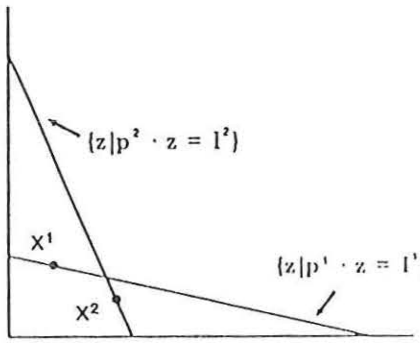

Figure 3

Finally, the equivalence between conditions (i) and (iv) implies that if one can find a nonsatiated function that is maximized at the observed x^i's then one can also find a monotone increasing, concave, and continuous function that is maximized at the observed x^i's.

Varian (1982) stated an alternative algebraic condition to Cyclical Consistency and developed algorithms to test the conditions of the above theorem.

Along similar lines to the above theorem, a large literature deals with nonparametric tests for the hypothesis that a given set of demand data has been generated from the maximization of a utility function that satisfies certain shape restrictions. For example, Afriat (1967b, 1972a, 1973, 1981), Diewert (1973), Diewert and Parkan (1985), and Varian (1983a) provided tests for the consistency of demand data with additively separable, weakly separable and homothetic utility functions. Matzkin and Richter (1991) provided a test for the strict concavity and strict monotonicity of the utility function; and Chiappori and Rochet (1987) developed a test for the consistency of demand data with a strictly concave and infinitely differentiable utility function. To provide an example of one such set of conditions, the algebraic conditions developed by Chiappori and Rochet are that (i) for all sequences $\{i, j, k, \ldots, r, l\}$ in $\{1, \ldots, N\}$

$$[p^j \cdot x^i \leqslant I^j, p^k \cdot x^j \leqslant I^k, \cdots, p^l \cdot x^r \leqslant I^l] \Rightarrow [I^i < p^i \cdot x^l], \text{ and}$$

(ii) for all i, j $[x^i = x^j] \Rightarrow [p^i = \alpha p^j$ for some $\alpha > 0]$.

Yatchew (1985) provided nonparametric restrictions for demand data generated by utility maximization subject to budget sets that are the union of linear sets. Matzkin (1991b) developed restrictions for demand data generated subject to choice sets that possess monotone and convex complement and for choices that are each supported by a unique hyperplane.

Nonstatistical nonparametric tests for the hypothesis of cost minimization and profit maximization have also been developed. See, for example, Afriat (1972b), Diewert and Parkan (1979), Hanoch and Rothschild (1978), Richter (1985) and Varian (1984). Suppose, for example, that $\{y^i, p^i\}$ are a set of observations on a vector of inputs and outputs, y, and a vector of the corresponding prices, P. Then, one of the results in the above papers is that $\{y^i, p^i\}$ is consistent with profit maximization iff for all $i,j = 1,\dots,N$, $p^i \cdot y^i \geqslant p^i \cdot y^j$ (Hanoch and Rothschild (1978)).

Some of the above mentioned tests have been used in empirical applications. See, for example, Landsburg (1981), McDonald and Manser (1984) and Manser and McDonald (1988).

Nonparametric restrictions have also been developed to test efficiency in production. These tests, typically appearing under the heading of Data Envelope Analysis, use data on the input and output vectors of different facilities (decision making units or DMU's) that are assumed to possess the same technology. Then, making assumptions about the technology, such as constant returns to scale, they determine the set of vectors of inputs and outputs that are efficient. A DMU is not efficient if its vector of input and output quantities is not in the efficiency set. See the paper by Seiford and Thrall (1990) for a survey of this literature.

Recently, nonparametric restrictions characterizing data generated by models other than the single agent optimization problem have been developed. Chiappori (1988) developed a test for the Pareto optimality of the consumption allocation within a household using data on aggregate household consumption and labor supply of each household member. Brown and Matzkin (1993) developed a test for the general equilibrium model, using data on market prices, aggregate endowments, consumers' incomes, and consumers' shares of profits. Nonparametric restrictions characterizing data consistent with other equilibrium models, such as various imperfect competition models, have not yet been developed. Varian (1983b) developed a test for the model of investors' behavior.

Some papers have developed statistical tests using the nonstatistical restrictions of some of the tests mentioned above (Varian (1985, 1988), Epstein and Yatchew (1985), Yatchew and Bos (1992) and Brown and Matzkin (1992), among others). As we will see in the next subsection, the test developed by Yatchew and Bos (1992), in particular, can be used with several of the above restrictions to obtain statistical nonparametric tests for economic models.

4.2. *Statistical tests*

Using nonparametric methods similar to those used to estimate nonparametric functions, it is possible to develop tests for the hypothesis that a nonparametric regression function satisfies a specified set of nonparametric shape restrictions. Yatchew and Bos (1992) and Gallant (1982), for example, present such tests.

The consistent test by Yatchew and Bos is based on a comparison of the restricted and unrestricted weighted sum of square errors. More specifically, suppose that the model is specified by $y = f^*(x) + \varepsilon$, where $y \in \mathbb{R}^q$, $x \in \mathbb{R}^K$, $\varepsilon \in \mathbb{R}^q$, x and ε are independent, $E(\varepsilon) = 0$, and $\mathrm{Cov}(\varepsilon) = \Sigma$. The null hypothesis is that $f^* \in \underline{\mathrm{F}} \subset \mathrm{F}$, while the alternative hypothesis is that $f^* \in \mathrm{F} \backslash \underline{\mathrm{F}}$. The Sobolev[6] norms of the functions in the sets F and $\underline{\mathrm{F}}$ are uniformly bounded. The test proceeds as follows. First, divide the sample into two independent samples of the same size, T. Compute the estimators $\underline{s}_T^2$ and s_T^2 using, respectively, the first and second samples, where

$$\underline{s}_T^2 = \min_{f \in \underline{\mathrm{F}}} \frac{1}{T} \Sigma_i [y^i - f(x^i)]' \Sigma^{-1} [y^i - f(x^i)]$$

and

$$s_T^2 = \min_{f \in \mathrm{F}} \frac{1}{T} \Sigma_i [y^i - f(x^i)]' \Sigma^{-1} [y^i - f(x^i)].$$

To transform these minimization problems into finite dimensional problems, Yatchew and Bos (1992) use a method similar to the one described in Section 3.1. They show that, under the null hypothesis, the asymptotic distribution of $t_{\mathrm{F}} = T^{1/2}[\underline{s}_T^2 - s_T^2]$ is $N(0, 2\upsilon)$, where $\upsilon = \mathrm{Var}\, \varepsilon' \Sigma^{-1} \varepsilon$. So, one can use standard statistical tables to determine whether the difference of the sum of squared errors is significantly different from zero. (This test builds on the work of Epstein and Yatchew (1985), Varian (1985) and Yatchew (1992).)

The Yatchew and Bos (1992) test can be used in conjunction with the nonstatistical nonparametric tests described in the previous subsection. Suppose for example that y^i denotes a vector of commodities purchased by a consumer and x^i denotes the vector of prices p^i and income I^i faced by the consumer when he or she purchased y^i. Assume that the observations are independent and for each i, $y^i = f^*(x^i) + \varepsilon$, where ε satisfies the assumptions made above. Then, as it is described in Yatchew and Bos (1992), we can use their method to test whether the data is consistent with the utility maximization hypothesis. In particular, Afriat's inequalities (in condition (iii) in Afriat's Theorem) can be used to calculate $\underline{s}_T^2$ by minimizing the value of

$$\sum_{i=1}^{T} [y^i - f^i]' \Sigma^{-1} [y^i - f^i]$$

with respect to V^i, λ^i, and f^i $(i = 1, \ldots, T)$ subject to (i) the Afriat inequalities: $V^i \leqslant V^j + \lambda^j p^j \cdot (f^i - f^j)$ $(i, j = 1, \ldots, T)$, (ii) the budget constraints: $p^i \cdot f^i = I^i$

[6] The Sobolev norm is defined on a set of m times continuously differentiable functions C^m by

$$\|f\| = \sum_{k=1}^{q} \left[\sum_{|\alpha| \leqslant m} \int [\mathrm{D}^\alpha f_k]^2 \,\mathrm{d}x \right]^{1/2}$$

where $\alpha = (\alpha_1, \ldots, \alpha_K)$ is a vector of integers; $\mathrm{D}^\alpha f(x)$ is the value resulting from differentiating f at x, α_1 times with respect to x_1, α_2 times with respect to $x_2, \ldots, \alpha_K$ times with respect to x_K; and $|\alpha| = \max_i |\alpha_i|$.

($i = 1, \ldots, T$), and (iii) inequalities that guarantee that the Sobolev norm of the function f is within specified bounds.

Gallant (1982) presents a seminonparametric method for testing whether a regression function satisfies some shape restrictions, such as linear homogeneity, separability and homotheticity. The method proceeds by testing whether the parametric approximation used to estimate a nonparametric function satisfies the hypothesized restrictions.

Following Gallant (1982), suppose that we are interested in testing the linear homogeneity of a cost function, $c(p, u)$, where $p = (p_1, \ldots, p_k)'$ is a vector of input prices and u is the output. Let

$$g(l, v) = \ln c\left(\frac{\exp l_1}{a_1}, \ldots, \frac{\exp l_K}{a_K}, \frac{\exp v}{a_{K+1}}\right),$$

where $l_i = \ln p_i + \ln a_i$ and $v = \ln u + \ln a_{K+1}$. (The a_i's are location parameters that are determined from the data.) Then, linear homogeneity of the cost function c in prices is equivalent to requiring that for all τ, $g(l + \tau 1, v) = \tau + g(l, v)$. The approximation g_N of g, given by

$$g_N(x \mid \theta) = b'x + x'Cx + \sum_{|k|^* \leq T} a_k \mathrm{e}^{\mathrm{i}k'x} \qquad x \in \mathbb{R}^{K+1}$$

satisfies these restrictions, for $C = \Sigma_k c_k k k'$, if

$$\sum_{j=1}^{K} b_j = 1 \quad \text{and if} \quad a_k = 0 \quad \text{and } c_k = 0 \quad \text{when} \quad \sum_{j=1}^{K} k_j \neq 0.$$

Linear homogeneity is then tested by determining whether these restrictions are satisfied. Gallant (1982) shows that by increasing the degree of approximation (i.e. the number of parameters) at a particular specified rate, as the number of observations increases, one can construct tests that are asymptotically free of specification bias. That is, for any given level of significance, α, one can construct a test statistic t_N and a critical value c_N such that if the true nonparametric function satisfies the null hypothesis then $\lim_{N \to \infty} \Pr(t_N > c_N) = \alpha$.

Several other methods have been developed to test restrictions of economic theory on nonparametric functions. Stoker (1989), for example, presents nonparametric tests for additive constraints on the first and second derivatives of a conditional mean function $f^*(x)$. These tests are based on weighted-average derivatives estimators (Stoker (1986), Powell et al. (1989), Härdle and Stoker (1989)). Linear homogeneity and symmetry constraints are examples of properties of f^* that can be tested using this method. (See also Lewbel (1991).) Also using average derivatives, Härdle et al. (1992) tested the positive definiteness of the matrix of aggregate income effects.

Hausman and Newey (1992) developed a test for the symmetry and negative slope of the Hicksian (compensated) demand. The test is derived from a nonparametric estimator for a consumer surplus. Since symmetry of the Hicksian demand implies that the consumer surplus is independent of the price path used to calculate it, estimates obtained using different paths should converge to the same limit. A minimum chi-square test is then developed using this idea.

We should also mention in this section the extensive existent literature that deals with tests for the monotonicity of nonparametric functions in a wide variety of statistical models. For a survey of such literature, we refer the reader to the previously mentioned books of Barlow et al. (1972) and Robertson et al. (1988). (See also Prakasa Rao (1983).)

5. Conclusion

We have discussed the use of restrictions implied by economic theory in the econometric analysis of nonparametric models. We described advancements that have been made on the theories of identification, estimation, and testing of nonparametric models due to the use of restrictions of economic theory.

First, we showed how restrictions implied by economic theory, such as shape and exclusion restrictions, can be used to identify functions in economic models. We demonstrated this in generalized regression models, binary threshold models, discrete choice models, models of consumer demand and in systems of simultaneous equations.

Various ways of incorporating economic shape restrictions into nonparametric estimators were discussed. Special attention was given to estimators whose feasibility depends critically on the imposition of shape restrictions. We described technical results that can be used to develop new shape restricted nonparametric estimators in a wide range of models. We also described seminonparametric and weighted average estimators and showed how one can impose restrictions of economic theory on estimators obtained by these two methods.

Finally, we have discussed some nonstatistical and statistical nonparametric tests. The nonstatistical tests are extensions of the basic ideas underlying the theory of Revealed Preference. The statistical tests are developed using nonparametric estimation methods.

References

Afriat, S. (1967a) "The Construction of a Utility Function from Demand Data", *International Economic Review*, 8, 66–77.

Afriat, S. (1967b) "The Construction of Separable Utility Functions from Expenditure Data", mimeo, Purdue University.

Afriat, S. (1972a) "The Theory of International Comparisons of Real Income and Prices", in: J.D. Daly,

ed., *International Comparisons of Prices and Output.* New York, National Bureau of Economic Research.
Afriat, S. (1972b) "Efficiency Estimates of Production Functions", *International Economic Review*, 13, 568–598.
Afriat, S. (1973) "On a System of Inequalities on Demand Analysis", *International Economic Review*, 14, 460–472.
Afriat, S. (1981) "On the Constructability of Consistent Price Indices between Several Periods Simultaneously", in: A. Deaton, ed., *Essays in Applied Demand Analysis.* Cambridge University Press: Cambridge.
Andrews, D.W.K. (1991) "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models", *Econometrica*, 59(2), 307–345.
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.
Balls, K.G. (1987) Inequality Constrained Nonparametric Estimation, Ph.D. Dissertation, Carnegie Mellon University.
Barlow, R.E., D.J. Bartholomew, J.M. Bremner and H.D. Brunk (1972) *Statistical Inference under Order Restrictions*, New York: John Wiley.
Barnett, W.A. and P. Yue (1988a) "The Asymptotically Ideal Model (AIM)", working paper.
Barnett, W.A. and P. Yue (1988b) "Semiparametric Estimation of the Asymptotically Ideal Model: The AIM Demand System", in: G. Rhodes and T. Fornby, eds., *Nonparametric and Robust Inference, Advances in Econometrics, Vol.* 7. JAI Press: Greenwich, Connecticut, 229–252.
Barnett, W.A., J. Geweke, and P. Yue (1991) "Seminonparametric Bayesian Estimation of the Asymptotically Ideal Model: The AIM Consumer Demand System", in: W. Barnett, J. Powell and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics.* Cambridge: Cambridge University Press.
Bergstrom, A.R. (1985) "The Estimation of Nonparametric Functions in a Hilbert Space", *Econometric Theory*, 1, 1–26.
Brown, D.J. and R.L. Matzkin (1991) "Recoverability and Estimation of the Demand and Utility Functions of Traders when Demands are Unobservable", mimeo, Cowles Foundation, Yale University.
Brown, D.J. and R.L. Matzkin (1992) "A Nonparametric Test for the Perfectly Competitive Model", mimeo, Northwestern University.
Brown, D.J. and R.L. Matzkin (1993) "Walrarian Comparative Statics", Technical Report No. 57, Stanford Institute for Theoretical Economics.
Brunk, H.D. (1970) "Estimation of Isotonic Regression", in: M.L. Puri, ed., *Nonparametric Techniques in Statistical Inference.* Cambridge University Press: Cambridge, 177–197.
Chiappori, P.A. (1988) "Rational Household Labor Supply", *Econometrica*, 56(1), 63–89.
Chiappori, P. and J. Rochet (1987) "Revealed Preference and Differential Demand", *Econometrica*, 55, 687–691.
Cosslett, S.R. (1983) "Distribution-free Maximum Likelihood Estimator of the Binary Choice Model", *Econometrica*, 51(3), 765–782.
Cox, D.R. (1970) *The Analysis of Binary Data.* Methuen & Co Ltd.
Diewert, E. (1973) "Afriat and Revealed Preference Theory", *Review of Economic Studies*, 40, 419–426.
Diewert, E.W. and C. Parkan (1979) "Linear Programming Tests for Regularity Conditions for Production Functions", University of British Columbia.
Diewert, E.W. and C. Parkan (1985) "Tests for the Consistency of Consumer Data", *Journal of Econometrics*, 30, 127–147.
Dykstra, R.L. (1983) "An Algorithm for Restricted Least Squares Regression", *Journal of the American Statistical Association*, 78, 837–842.
Eastwood, B.J. (1991) "Asymptotic Normality and Consistency of Semi-nonparametric Regression Estimators Using an upward F Test Truncation Rule", *Journal of Econometrics.*
Eastwood, B.J. and A.R. Gallant (1991) "Adaptive Truncation Rules for Seminonparametric Estimators that Achieve Asymptotic Normality", *Econometric Theory*, 7, 307–340.
Epstein, L.G. and D.J. Yatchew (1985) "Non-Parametric Hypothesis Testing Procedures and Applications to Demand Analysis", *Journal of Econometrics*, 30, 150–169.
Finney, D.J. (1971) *Probit Analysis*, Cambridge University Press.
Friedman, J. and R. Tibshirani (1984) "The Monotone Smoothing of Scatter Plots", *Technometrics*, 26, 243–350.

Gallant, A.R. (1981) "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form", *Journal of Econometrics*, 15, 211–245.
Gallant, A.R. (1982) "Unbiased Determination of Production Technologies", *Journal of Econometrics*, 20, 285–323.
Gallant, A.R. (1987) "Identification and Consistency in Seminonparametric Regression", in: T.F. Bewley, ed., *Advances in Econometrics, Fifth World Congress, Volume I*. Cambridge University Press.
Gallant, A.R. and G.H. Golub (1984) "Imposing Curvature Restrictions on Flexible Functional Forms", *Journal of Econometrics*, 26, 295–321.
Gallant, A.R. and D.W. Nychka (1987) "Seminonparametric Maximum Likelihood Estimation", *Econometrica*, 55, 363–390.
Gallant, A.R. and G. Souza (1991) "On the Asymptotic Normality of Fourier Flexible Form Estimates", *Journal of Econometrics*, 50, 329–353.
Goldman, S.M. and P.A. Ruud (1992) "Nonparametric Multivariate Regression Subject to Monotonicity and Convexity Constraints", mimeo, University of California, Berkeley.
Han, A.K. (1987) "Nonparametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimation", *Journal of Econometrics*, 35, 303–316.
Hanoch, G. and M. Rothschild (1978) "Testing the Assumptions of Production Theory: A Nonparametric Approach", *Journal of Political Economy*, 80, 256–275.
Hanson, D.L. and G. Pledger (1976) "Consistency in Concave Regression", *Annals of Statistics*, 4, 1038–1050.
Hanson, D.L., G. Pledger, and F.T. Wright (1973) "On Consistency in Monotonic Regression", *Annals of Statistics*, 1, 401–421.
Härdle, W. and T.M. Stoker (1989) "Investigating Smooth Multiple Regression by the Method of Average Derivatives", *Journal of the American Statistical Association*, 84, 986–995.
Härdle, W., W. Hildenbrand and M. Jerison (1992) "Empirical Evidence on the Law of Demand", *Econometrica*, 59, 1525–1550.
Hausman, J. and W. Newey (1992) "Nonparametric Estimation of Exact Consumer Surplus and Deadweight Loss", mimeo, Department of Economics, M.I.T.
Hausman, J.A. and D.A. Wise (1978) "A Conditional Probit Model of Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences", *Econometrica*, 46, 403–426.
Hildreth, C. (1954) "Point Estimates of Ordinates of Concave Functions", *Journal of the American Statistical Association*, 49, 598–619.
Horowitz, J.L. (1992) "A Smoothed Maximum Score Estimator for the Binary Choice Model", *Econometrica*, 60, 505–531.
Hotz, V.J. and R.A. Miller (1989) "Conditional Choice Probabilities and the Estimation of Dynamic Discrete Choice Models", mimeo, University of Chicago and Carnegie Mellon University.
Houthakker, H.S. (1950) "Revealed Preference and the Utility Function", *Economica*, 17, 159–174.
Ichimura, H. (1993) "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models", *Journal of Econometrics*, 58, 71–120.
Kadiyali, V. (1993) Ph.D. Thesis, Department of Economics, Northwestern University.
Kiefer, J. and J. Wolfowitz (1956) "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters", *Annals of Mathematical Statistics*, 27, 887–906.
Kimeldorf, G.S. and G. Wahba (1971) "Some Results on Tchebycheffian Splines Functions", *Journal of Mathematical Analysis and Applications*, 33, 82–95.
Klein, R.W. and R.H. Spady (1993) "An Efficient Semiparametric Estimator for Discrete Choice Models", *Econometrica*, 61, 387–422.
Landsburg, S.E. (1981) "Taste Change in the United Kingdom, 1900–1955", *Journal of Political Economy*, 89, 92–104.
Lewbel, A. (1991) "Applied Consistent Tests of Nonparametric Regression and Density Restriction", mimeo, Brandeis University.
Maddala, G.S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press: Cambridge.
Mammen, E. (1991a) "Estimating a Smooth Monotone Function", *Annals of Statistics*, 19, 724–740.
Mammen, E. (1991b) "Nonparametric Regression under Qualitative Smoothness Assumptions", *Annals of Statistics*, 19, 741–759.
Manser, M.E. and R.J. McDonald (1988) "An Analysis of Substitution Bias in Measuring Inflation, 1959–85", *Econometrica*, 56(4), 909–930.

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

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

Manski, C. (1988) "Identification of Binary Response Models", *Journal of the American Statistical Association*, 83, 729–738.

Mas-Colell, A. (1977) "On the Recoverability of Consumers' Preferences from Market Demand Behavior", *Econometrica*, 45(6), 1409–1430.

Mas-Colell, A. (1978) "On Revealed Preference Analysis", *Review of Economic Studies*, 45, 121–131.

Matzkin, R.L. (1986) Mathematical and Statistical Inferences from Demand Data, Ph.D. Dissertation, University of Minnesota.

Matzkin, R.L. (1990a) "Estimation of Multinomial Models Using Weak Monotonicity Assumptions", Cowles Foundation Discussion Paper No. 957, Yale University.

Matzkin, R.L. (1990b) "Least-concavity and the Distribution-free Estimation of Nonparametric Concave Functions", Cowles Foundation Discussion Paper No. 958, Yale University.

Matzkin, R.L. (1990c) "Fully Nonparametric Estimation of Some Qualitative Dependent Variable Models Using the Method of Kernels", mimeo, Cowles Foundation, Yale University.

Matzkin, R.L. (1991a) "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models", *Econometrica*, 59, 1315–1327.

Matzkin, R.L. (1991b) "Axioms of Revealed Preference for Nonlinear Choice Sets", *Econometrica*, 59, 1779–1786.

Matzkin, R.L. (1991c) "A Nonparametric Maximum Rank Correlation Estimator", in: W. Barnett, J. Powell and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge: Cambridge University Press.

Matzkin, R.L. (1991d) "Using Kernel Methods to Smooth Concavity Restricted Estimators", mimeo, Cowles Foundation, Yale University.

Matzkin, R.L. (1992) "Nonparametric and Distribution-Free Estimation of the Binary Choice and the Threshold-Crossing Models", *Econometrica*, 60, 239–270.

Matzkin, R.L. (1993a) "Nonparametric Identification and Estimation of Polychotomous Choice Models", *Journal of Econometrics*, 58, 137–168.

Matzkin, R.L. (1993b) "Computation and Operational Properties of Nonparametric Concavity Restricted Estimators", Northwestern University.

Matzkin, R.L. (1994) "Identification in Nonparametric LDV Models", mimeo, Northwestern University.

Matzkin, R.L. and W. Newey (1992) "Kernel Estimation of a Structural Nonparametric Limited Dependent Variable Models", mimeo.

Matzkin, R.L. and M.K. Richter (1991) "Testing Strictly Concave Rationality", *Journal of Economic Theory*, 53, 287–303.

McDonald, R.J. and M.E. Manser (1984) "The Effect of Commodity Aggregation on Tests of Consumer Behavior", mimeo.

McFadden, D. (1974) "Conditional Logit Analysis of Qualitative Choice Behavior", in P. Zarembka, ed., *Frontiers of Econometrics*, New York: Academic Press, 105–142.

McFadden, D. (1981) "Econometric Models of Probabilistic Choice", in: C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*. The MIT Press, 198–272.

McFadden, D. (1985) "Specification of Econometric Models", Presidential Address, Fifth World Congress, mimeo, M.I.T.

McFadden, D. and M.K. Richter (1970) "Stochastic Rationality and Revealed Stochastic Preference", Department of Economics, University of California, Berkeley, mimeo.

McFadden, D. and M.K. Richter (1990) "Stochastic Rationality and Revealed Stochastic Preference", in: J.S. Chipman, D. McFadden and M.K. Richter, eds., *Preference, Uncertainty, and Optimality. Essays in Honor of Leonid Hurwicz*. Boulder: Westview Press, 161–186.

Mukarjee, H. (1988) "Monotone Nonparametric Regression", *The Annals of Statistics*, 16, 741–750.

Mukarjee, H. and S. Stern (1994) "Feasible Nonparametric Estimation of Multiargument Monotone Functions", *Journal of the American Statistical Association*, 89, No. 425, 77–80.

Nadaraja, E. (1964) "On Regression Estimators", *Theory of Probability and its Applications*, 9, 157–159.

Nemirovskii, A.S., B.T. Polyak, and A.B. Tsybakov (1983) "Rates of Convergence of Nonparametric Estimates of Maximum Likelihood Type", *Problem peredachi informatsii*, 21, 258–272.

Powell, J.L., J.H. Stock, and T.M. Stoker (1989), "Semiparametric Estimation of Index Coefficients", *Econometrica*, 57, 1403–1430.
Prakasa Rao, B.L.S. (1983) *Nonparametric Functional Estimation*, Academic Press.
Richter, M.K. (1985) "Theory of Profit", mimeo, Department of Economics, University of Minnesota.
Robertson, T., F.T. Wright, and R.L. Dykstra (1988) *Order Restricted Statistical Inference*, John Wiley and Sons.
Roehrig, C.S. (1988) "Conditions for Identification in Nonparametric and Parametric Models", *Econometrica*, 56(2), 433–447.
Royall, R.M. (1966) A Class of Nonparametric Estimators of a Smooth Regression Function. Ph.D. Thesis, Stanford University.
Samuelson, P.A. (1938) "A Note on the Pure Theory of Consumer Behavior", *Economica*, 5, 61–71.
Seiford, L.M. and R.M. Thrall (1990) "Recent Development in DEA", *Journal of Econometrics*, 46, 7–38.
Stoker, T.M. (1986) "Consistent Estimation of Scaled Coefficients", *Econometrica*, 54, 1461–1481.
Stoker, T.M. (1989) "Tests of Additive Derivatives Constraints", *Review of Economic Studies*, 56, 535–552.
Strauss, D. (1979) "Some Results on Random Utility Models", *Journal of Mathematical Psychology*, 20, 35–52.
Thompson, T.S. (1989) "Identification of Semiparametric Discrete Choice Models", Discussion Paper No. 249, Center for Economic Research, University of Minnesota.
Utreras, F. (1984) "Positive Thin Plate Splines", CAT Rep. No. 68, Dept. of Math., Texas A & M Univ.
Utreras, F. (1985) "Smoothing Noisy Data Under Monotonicity Constraints: Existence, Characterization, and Convergence Rates", *Numerische Mathematik*, 47, 611–625.
Varian, H. (1982) "The Nonparametric Approach to Demand Analysis", *Econometrica*, 50(4), 945–974.
Varian, H. (1983a) "Nonparametric Tests of Consumer Behavior", *Review of Economic Studies*, 50, 99–110.
Varian, H. (1983b) "Nonparametric Tests of Models of Investor Behavior", *Journal of Financial and Quantitative Analysis*, 18, 269–278.
Varian, H. (1984) "The Nonparametric Approach to Production Analysis", *Econometrica*, 52, 579–597.
Varian, H. (1985) "Non-Parametric Analysis of Optimizing Behavior with Measurement Error", *Journal of Econometrics*, 30.
Varian, H. (1988) "Goodness-of-Fit in Demand Analysis", CREST Working Paper No. 89–11, University of Michigan.
Villalobos, M. and G. Wahba (1987) "Inequality Constrained Multivariate Smoothing Splines with Application to the Estimation of Posterior Probabilities", *Journal of the American Statistical Association*, 82, 239–248.
Wahba, G. (1990) *Spline Models for Observational Data*, CBMS–NSG Regional Conference Series in Applied Mathematics, No. 59, Society for Industrial and Applied Mathematics.
Wald, A. (1949) "Note on the Consistency of the Maximum Likelihood Estimator", *Annals of Mathematical Statistics*, 20, 595–601.
Wang, Y. (1992) Nonparametric Estimation Subject to Shape Restrictions, Ph.D. Dissertation, Department of Statistics, University of California, Berkeley.
Watson, G.S. (1964) "Smooth Regression Analysis", *Sankhya Series A*, 26, 359–372.
Wong, W.H. (1984) "On Constrained Multivariate Splines and Their Approximations", *Numerische Mathematik*, 43, 141–152.
Wright, F.T. (1982) "Monotone Regression Estimates for Grouped Observations", *Annals of Statistics*, 10, 278–286.
Yatchew, A.J. (1985) "A Note on Nonparametric Tests of Consumer Behavior", *Economic Letters*, 18, 45–48.
Yatchew, A.J. (1992) "Nonparametric Regression Tests Based on Least Squares", *Econometric Theory*, 8, 435–451.
Yatchew, A. and L. Bos (1992) "Nonparametric Tests of Demand Theory", mimeo, Department of Economics, University of Toronto.
Yellott, J.I. (1977) "The Relationship Between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgement, and the Double Exponential Distribution", *Journal of Mathematical Psychology*, 15, 109–144.

NONPARAMETRIC DISCRETE CHOICE MODELS
WITH UNOBSERVED HETEROGENEITY

Richard Briesch
New York University

Pradeep Chintagunta
University of Chicago

Rosa Matzkin
Northwestern University

August 1996

ABSTRACT

We introduce a nonparametric method for estimating discrete choice models that does not require that different consumers possess the same systematic subutility. The method lessens the possibility of misspecification because it does not require specifying that the unknown functions and distributions in the model belong to known parametric families. Neither the systematic subutilities, nor the distribution of the random subutilitites, nor the distribution of the systematic subutilities over the population is required to be specified parametrically. We show that the estimators that we propose for these functions and distributions are strongly consistent.

The method is an extension of the fully nonparametric estimators for discrete choice models introduced in Matzkin (1992, 1993a), where the subutility function and the distribution of the random term were nonparametric but all the consumers were assumed to possess the same subutility function.

The estimators are obtained by maximizing a log-likelihood function over a set of functions and distributions. We show that these estimators can be computed by transforming this maximization problem into the maximization of a finite dimensional vector subject to inequality constraints.

* Rosa L. Matzkin gratefully acknowledges support for this research from the National Science Foundation.

1. INTRODUCTION

Discrete choice models are one of the most popular models that are used to analyze the choice of a typical consumer between a finite number of alternatives. These models have been applied to analyze, among others, the choice between various means of transportation, the choice between different schools, and the choice between different brands of a product. In this model, each alternative is characterized by a vector of attributes, and the consumer chooses the alternative from which he/she derives the highest level of utility. The utility is the sum of a subutility of observed attributes - the "systematic subutility" - and a random term - the unobserved "random subutility". For example, in a model of a commuter choosing between various means of transportation, the alternatives may be car and bus, and the observable attributes may be the cost and time associated with each alternative. The utility of the commuter for a means of transportation is the sum of a function of the time and cost of that means of transportation (the systematic subutility) and an unobservable random term that represents the value of a subutility of unobserved attributes of the means of transportation, such as comfort.

For the most part, the estimation of discrete choice models has proceeded in the past by assuming that the systematic subutility function and the distribution of the unobservable random subutilities are the same for all consumers and are known up to a finite dimensional vector. Typically, the systematic subutility has been assumed to be linear in the parameters and the distribution of the random term has been specified to be either a Weibull or a normal distribution. Several potential problems can be caused, however, when using this type of assumptions. First, the specifications chosen for the systematic subutility function and for the distribution of the unobservable random terms may be incorrect, yielding the corresponding estimators to be, in general, inconsistent. Second, different consumers may posses different systematic subutility and distribution functions, even after controlling for observed characteristics. If this "unobserved heterogeneity" is not taken into consideration, the estimators will also be, in general, inconsistent.

To avoid estimating the distribution and/or the systematic subutility function inconsistently due to parametric misspecifications, several methods have been developed, following the pioneering work of Manski (1975). Manski (1975, 1985), Cosslett (1983), \ Klein and Spady (1993), and Horowitz (1992), among others, developed methods where the distribution of the random term is not specified parametrically, while Matzkin (1991), among others, developed methods where the systematic subutility function does not need to be specified parametrically. The methods in Manski (1975,1985) and Horowitz (1992) allow the distribution of the unobservable additive random terms to be different for different consumers, as long as these consumers possess different values for the observable exogenous variables. Methods where neither the systematic subutility function nor the distribution of the random term are specified parametrically were introduced in Matzkin (1992, 1993). All these methods, however, assume that there is no unobserved heterogeneity in the systematic subutility function. In other words, all these methods assume that all the consumers have the same systematic subutility function, given the observable exogenous variables.

The importance of unobserved heterogeneity has been recognized for a long time. Heckman (1974) and Heckman and Willis (1977) allow for heterogeneity in preferences, and Heckman and Singer (1984) introduced a method of estimating the distribution of an unobserved heterogeneity parameter without specifying a parametric distribution for this parameter. This latter paper also documents the serious inconsistencies that may arise when unobserved heterogeneity is not taken into consideration when estimating a model. (See also Heckman (1981a, 1981b) and Heckman and Walker (1990a, 1990b).) A model of Thurstone's (1930) deals with heterogeneity of preferences in discrete choice models, and more recently, Albright, Lerman, and Manski (1977), Hausman and Wise (1978), and Ichimura and Thompson (1993) have presented methods to estimate discrete choice models with heterogeneous preferences. These methods require, however, that at least some of the functions and distributions in the model be known up to the value of a finite dimensional parameter. Albright, Lerman, and Manski (1977) and Hausman and Wise (1978) specify all the functions and distributions in the model up to a finite dimensional parameter. Ichimura and Thompson (1993) specify the systematic subutility parametrically, as a linear function of random parameters.

In this paper, we present a nonparametric estimation method for discrete choice models that allows the systematic subutility to be different across individuals and that requires fewer parametric assumptions than any of the currently available methods. Neither the systematic subutility, nor the distribution of the unobservable random terms, nor the distribution of the systematic subutility over the population is assumed in our method to be completely parametrically specified. We propose a method for estimating all these functions and distributions consistently.

In the next section we describe the model. In Section 3 we state conditions under which the model is identified. In Section 4 we present strongly consistent estimators for the functions and distributions in the model. A method to compute the estimators is presented in Section 5; and a brief conclusion is presented in Section 6.

2. THE MODEL:

As usual in discrete choice models, we assume that a typical consumer must choose one of a finite number, J, of alternatives, and he/she chooses the one that maximizes the value of a utility function that depends on the characteristics of the alternatives and the consumer.. Each alternative j is characterized by a vector, z_j, of the observable attributes of the alternatives. We will assume that $z_j \equiv (x_j, r_j)$, where $r_j \in R$ and $x_j \in R^K$ $(K \geq 1)$. Each consumer is characterized by a vector, $s \in R^L$, of observable socioeconomic characteristics for the consumer. The utility of a consumer with observable socioeconomic characteristics s, for an alternative, j, is given by

$$V^*(j, s, z_j, \omega) + \epsilon_j$$

where ϵ_j and ω denote the values of unobservable random variables. For any given value of ω, and any j, $V^*(j, \cdot, \omega)$ is a real valued, but otherwise unknown, function. The dependence of V^* on ω allows us to incorporate into the model the possibility that this systematic subutility be different from different consumers, even if the observable exogenous variables posses the same values for these consumers. We will denote the distribution of ω by G^* and we will denote the distribution of the random vector $(\epsilon_1, ..., \epsilon_J)$ by F^*.

The probability that a consumer with socioeconomic characteristics s will choose alternative j when the vector of observable attributes of the alternatives is $z \equiv (z_1, ..., z_J) \equiv (x_1, r_1, ..., x_J, r_J)$ will be denoted by $p(j|s, z; V^*, F^*, G^*)$. Hence,

$$p(j|s, z; V, F, G) = \int \Pr(j|s, z; \omega, V, F) \ dG(\omega),$$

where $\Pr(j|s, z; \omega, V, F)$ denotes the probability that a consumer with systematic subutility $V(\cdot; \omega)$ will choose alternative j, when the distribution of ϵ is F. By the utility maximization hypothesis,

$$\Pr(j|s, z; \omega, V, F)$$

$$= \Pr\{V(j, s, x_j, r_j, \omega) + \epsilon_j > V(k, s, x_k, w_k, \omega) + \epsilon_k \text{ for all } k \neq j\}$$

$$= \Pr\left\{\epsilon_k - \epsilon_j < V(j,s,x_j,r_j,\omega) - V(k,s,x_k,w_k,\omega) \text{ for all } k \neq j\right\}$$

which depends on the distribution F. In particular, if we let F_1 denote the distribution of the vector $(\epsilon_2 - \epsilon_1, ..., \epsilon_J - \epsilon_1)$, then

$$\Pr(1|s,z;\omega,V,F)$$

$$= F_1\left(V(1,s,x_1,w_1,\omega) - V(2,s,x_2,w_2,\omega), ..., V(1,s,x_1,w_1,\omega) - V(J,s,x_2,w_2,\omega)\right)$$

and the probability that the consumer will choose alternative 1 is then

$$p(1|s,z;V,F,G) =$$

$$\int \Pr(1|s,z;\omega,V,F)\ dG(\omega),$$

$$= \int F_1\left(V(1,s,x_1,w_1,\omega) - V(2,s,x_1,w_1,\omega), ..., V(1,s,x_1,w_1,\omega) - V(J,s,x_1,w_1,\omega)\right)\ dG(\omega)$$

For any j, $\Pr(j|s,z;\omega,V,F)$ can be obtained in an analogous way, letting F_j denote the distribution of $(\epsilon_1 - \epsilon_j, ..., \epsilon_J - \epsilon_j)$.

3. NONPARAMETRIC IDENTIFICATION

Our objective is to develop estimators for the function V^* and the distributions F^* and G^*, without requiring that these functions and distributions belong to parametric families. It follows from the definition of the model that we can only hope to identify the distributions of the vectors $\eta_j \equiv (\epsilon_1 - \epsilon_j, ..., \epsilon_J - \epsilon_j)$ for $j = 1, ..., J$. Let F_j^* denote the distribution of η_j $(j = 1, ..., J)$. We will denote by $\underline{F}^*$ the vector $(F_1^*, ..., F_J^*)$.

Adapting the standard definition of identification to our model, we can state the following:

DEFINITION: The function V^* and the distributions $\underline{F}^*$ and G^* are identified in a set $(W \times \Gamma_F \times \Gamma_G)$ such that $(V^*, \underline{F}^*, G^*) \in (W \times \Gamma_F \times \Gamma_G)$ if $\forall (V, \underline{F}, G) \in$

$(W \times \Gamma_F \times \Gamma_G)$

$$\int \Pr(j|s,z;V(\cdot;\omega),F)\, dG(\omega) = \int \Pr(j|s,z;V^*(\cdot;\omega),F^*)\, dG^*(\omega)$$

$$for \;\; j = 1,\ldots,J$$

$$\Rightarrow V = V^*,\ \underline{F} = \underline{F}^* \;\; \& \; G = G^*.$$

That is, $(V^*, \underline{F}^*, G^*)$ is identified in a set $(W \times \Gamma_F \times \Gamma_G)$ to which $(V^*, \underline{F}^*, G^*)$ belongs, if any triple, $(V, \underline{F}, G)$, that belongs to $(W \times \Gamma_F \times \Gamma_G)$ and is different from $(V^*, \underline{F}^*, G^*)$ generates, for at least one alternative j, choice probabilities, $p(j|s,z;V,\underline{F},G)$, that are different (a.s.) from $p(1|s,z;V^*,F^*,G^*)$.

We next present a set of conditions that, when satisfied, guarantee that $(V^*, \underline{F}^*, G^*)$ is identified in $(W \times \Gamma_F \times \Gamma_G)$.

ASSUMPTION 0: W is a set of differentiable functions, and Γ_G and Γ_F are sets of absolutely continuous distributions.

ASSUMPTION 1: For all $j \in \{1,\ldots,J\}$, the random vector $\eta_j \equiv (\epsilon_1 - \epsilon_j,\ldots,\epsilon_J - \epsilon_j)$ is distributed independently of $(s, z_1, \ldots, z_J)$ and ω.

ASSUMPTION 2: The random variable ω is distributed independently of $(s, z_1, \ldots, z_J)$.

ASSUMPTION 3: The support of the probability measure of ω is R.

ASSUMPTION 4: For all j, ω, and $V \in W$ there exists a real valued function $v(j,\cdot,\cdot,\omega)$ such that $\forall (s, x_j, r_j) \;\; V(j,s,x_j,r_j,\omega) = v(j,s,x_j,\omega) + r_j$.

ASSUMPTION 5: $\exists\ \bar{s}, \bar{x}_1, \ldots, \bar{x}_J$ and $(\alpha_1,\ldots,\alpha_J) \in R^J$ such that $\forall j\ \forall \omega\ \forall V \in W$, $v(j,\bar{s},\bar{x}_j,\omega) = \alpha_j$.

ASSUMPTION 6: $\exists \tilde{j}$, $\beta_j \in R$, and $\tilde{x}_{\tilde{j}}$ such that $\forall s\ \forall \omega\ \forall V \in W$, $v(j,s,\tilde{x}_{\tilde{j}},\omega) = \beta_{\tilde{j}}$.

ASSUMPTION 7: $\exists j^* \neq \tilde{j}$ such that $\forall v(j^*,\cdot) \neq v'(j^*\cdot\cdot)$ $\exists \hat{s}, \hat{x}, \hat{\omega}, \hat{\omega}'$ such that $v(j^*, \hat{s}, \hat{x}, \hat{\omega}) = v'(j^*, \hat{s}, \hat{x}, \hat{\omega}')$ and $\frac{\partial v(j^*,\hat{s},\hat{x}_j,\bullet,\hat{\omega})}{\partial(s,x_j,\bullet)} \neq \frac{\partial v'(j^*,\hat{s},\hat{x}_j,\bullet,\hat{\omega}')}{\partial(s,x_j,\bullet)}$.

ASSUMPTION 8: One of the following holds:

(8.i) $\forall j \neq j^*, \tilde{j}$, $\forall v, v'$ such that $v(j,\cdot) \neq v^*(j,\cdot)$ $\exists s, x, \omega, \omega'$ such that $v(j,s,x,\omega) - v'(j,s,x,\omega')$ and $\frac{\partial v(j,s,x_j,\omega)}{\partial(s,x_j)} \neq \frac{\partial v'(j,s,x_j,\omega')}{\partial\partial(s,x_j)}$

and

$\exists \hat{j} \neq \tilde{j}$ $\beta_{\hat{j}} \in R$, and $\hat{x}_{\hat{j}}$ such that $\forall s$ $\forall \omega$ $\forall V \in W$, $v(\hat{j}, s, \hat{x}_{\hat{j}}, \omega) = \beta_{\hat{j}}$.

(8.ii) $\forall j \neq j^*, \tilde{j}$, $\forall v, v'$ such that $v(j,\cdot) \neq v^*(j,\cdot)$ $\exists s, x, \omega, \omega'$ such that $v^*(j,s,x,\omega) - v'(j,s,x,\omega')$ and $\frac{\partial v(j,s,x_j,\omega)}{\partial(s,x_j)} \neq \frac{\partial v'(j,s,x_j,\omega')}{\partial\partial(s,x_j)}$

and

$\exists k$ such that $\forall s, x_{\tilde{j}}, x_k$ $\forall v$, the function $n_{\tilde{j},k}(s, x_{\tilde{j}}, x_k, \omega) \equiv v(\tilde{j}, s, x_{\tilde{j}}, \omega) - v(k, s, x_k, \omega)$ is either strictly increasing or strictly decreasing in ω.

(8.iii) $\forall j$ it is possible to find a sequence $\{j_1, ..., j_M\}$ of alternatives such that $j_1 = j$, $j_M = j^*$ or $\tilde{j}$, and $\forall k \in \{1, ..., M-1\}$ the function $n_{k,k+1}(s, x_k, x_{k+1}, \omega) \equiv v(k, s, x_k, \omega) - v(k+1, s, x_{k+1}, \omega)$ is either strictly increasing or strictly decreasing in ω.

ASSUMPTION 9: $\forall j$, the support of r_j is R.

ASSUMPTION 10 : G^* is strictly increasing.

Assumption 0 implies that when $(V^*, \underline{F}^*, G^*) \in (W \times \Gamma_F \times \Gamma_G)$, V^* is a differentiable function, and $\underline{F}^*$ and G^* are absolutely continuous distributions. Assumption 1, together with Assumptions 4 and 5, allow us to identify $\underline{F}^*$. To see this, note that Assumption 4 states that the systematic subutility is additive separable into the value of the observable variable r_j and a function of all the other arguments, and Assumption 5 states that the values of these latter functions are fixed at one point $(\bar{s}, \bar{x}_1, ..., \bar{x}_J)$. Hence, the variation in the values of $p(1|\bar{s}, \bar{x}_1, r_1, ..., \bar{x}_J, r_J; V^*, \underline{F}^*, G^*)$ over different values of $(r_1, ..., r_J)$ can only be attributed to $\underline{F}^*$ .The differentiability of V^* , the absolute continuity of G^*, and Assumptions 2, 3, 6 and 7 allow us to use a result in Brown and Matzkin (1995) to show that G^* and a function $V^*(j, \cdot, \omega)$ are identified. Assumption 8 allow us to use F^* and G^* to identify the differences $V^*(j, \cdot, \omega) - V^*(k, \cdot, \omega)$. Assumption 6 allows us to identify each $V^*(j, \cdot, \omega)$ function from these differences. Assumption 7 guarantees that the support of the difference $r_j - r_k$, for all j, k, is the real line and that the value of r_j can be observed at large negative values. The first property guarantees that the values of F_j^* can be recovered over the whole R^{J-1} space. The second property guarantees that in the limit, when for all but two $j's$, $r_j \to -\infty$, the observed values of $p(1|s, z; V^*, F^*, G^*)$ correspond to the values of a binary choice model.

In the following examples, Assumptions 5-8 are satisfied:

EXAMPLE 1: A binary choice model where each function $v(1, \cdot)$ is characterized by a function $r(\cdot)$, and each function $v(2, \cdot)$ is characterized by a function $h(\cdot)$ such that for all s, x_1, x_2, ω, where $x_1 = (x_1^{(1)}, x_1^{(2)})$ and $x_1^{(2)} \in R$:

(i) $v(1, s, x_1, \omega) = r(s, x_1^{(1)}) + \omega\, x_1^{(2)}$

(ii) $v(2, s, x_2, \omega) = h(x_2, \omega)$

(iii) $r(0, 0) = 0$

(iv) $h(0, \omega) = 0$ for all ω

(v) the support of $x_1^{(2)}$ is R_+, and

(v) h is strictly decreasing in ω, for all x_2.

In this example, Assumption 5 is satisfied when $\bar{s} = 0$, $\bar{x}_1 = (\bar{x}_1^{(1)}, \bar{x}_1^{(2)}) = (0,0)$, $\bar{x}_2 = 0$, $\alpha_1 = \alpha_2 = 0$. Assumption 6 is satisfied for $\tilde{j} = 2$, when $\beta_{\tilde{j}} = 0$ and $\tilde{x}_2 = 0$. Assumption 7 is satisfied for $j^* = 1$. To see this, note that if $v(1,\cdot) \neq v'(1.\cdot)$, then it must be that for some vector $(\check{s}, \check{x}_1^{(1)})$, $r(\check{s}, \check{x}_1^{(1)}) \neq r'(\check{s}, \check{x}_1^{(1)})$, where $v(1,s,x_1,\omega) = r(s, x_1^{(1)}) + \omega\ x_1^{(2)}$ and $v'(1,s,x_1,\omega) = r'(s, x_1^{(1)}) + \omega\ x_1^{(2)}$. Hence, since $r(0,0) = r'(0,0)$, there must exist a vector $(\dot{s}, \dot{x}_1^{(1)})$, such that $\frac{\partial r(\dot{s},\dot{x}_1^{(1)})}{\partial(s,x_1)} \neq \frac{\partial r'(\dot{s},\dot{x}_1^{(1)})}{\partial(s,x_1)}$. Let $\dot{x}_1^{(2)}, \omega, \omega'$ be such that $r(\dot{s}, \dot{x}_1^{(1)}) + \omega\ \dot{x}_1^{(2)} = r'(\dot{s}, \dot{x}_1^{(1)}) + \omega\ \dot{x}_1^{(2)}$. Then, Assumption 7 is satisfied. Finally, to see that Assumption 8 is satisfied, note that for all $(s, x_1^{(1)}, x_1^{(2)}, x_2, \omega)$, $\frac{\partial h(x_2,\omega)}{\partial \omega} - x_1^{(2)} < 0$. It follows that when $j = 2$, $n_{2,1}(s,x_1,x_2,\omega) = v(2,s,x_2,\omega) - v(1,s,x_1,\omega)$ is strictly decreasing in ω, and when $j = 1$, $n_{1,2}(s,x_2,x_1,\omega) = v(1,s,x_1,\omega) - v(2,s,x_2,\omega)$ is strictly increasing in ω. Hence Assumption 8(iii) is satisfied.

EXAMPLE 2: A trichotomous choice model where $s \in R$, each function $v(1,\cdot)$ is characterized by a function $h(\cdot)$, each function $v(2,\cdot)$ is characterized by a function $t(\cdot)$, and each function $v(3,\cdot)$ is characterized by a function $m(\cdot)$, such that for all s, x_1, x_2, x_3, ω :

(i) $v(1,s,x_1,\omega) = \omega\ s\ + h(x_1)$

(ii) $v(2,s,x_2,\omega) = t(x_2,\omega)$

(iii) $v(3,s,x_3,\omega) = m(s,x_3,\omega)$

(iv) $h(0) = 0$

(v) $t(0,\omega) = 0$ for all ω

(vi) $m(0,0,\omega) = 0$ for all ω,

(v) the support of s is R_+, and

(v) m is strictly decreasing in ω, for all (s, x_3, ω).

In this example, Assumption 5 is satisfied when $\bar{s} = 0$, $\bar{x}_1 = 0$, $\bar{x}_2 = 0$, $\bar{x}_3 = 0$, and $\alpha_1 = \alpha_2 = 0$. Assumption 6 is satisfied for $\tilde{j} = 2$, when $\beta_{\tilde{j}} = 0$ and $\tilde{x}_2 = 0$. Assumption 7 is satisfied for $j^* = 1$, by the same argument used to show that this assumption was satisfied in the previous example. And Assumption 8 is satisfied because $n_{1,2}(s, x_2, x_1, \omega) = v(1, s, x_1, \omega) - v(2, s, x_2, \omega)$ is strictly increasing in ω, $n_{2,1}(s, x_1, x_2, \omega) = v(2, s, x_2, \omega) - v(1, s, x_1, \omega)$ is strictly decreasing in ω, and $n_{3,1}(s, x_3, x_1, \omega) = v(3, s, x_3, \omega) - v(1, s, x_1, \omega)$ is strictly decreasing in ω.

Using the assumptions specified above, we can prove the following theorem:

THEOREM 1: If assumptions 1-8 are satisfied, then $(V^*, G^*, \underline{F}^*)$ is identified in $(W \times \Gamma_F \times \Gamma_G)$.

The proof of the theorem is presented in the Appendix.

This theorem shows that one can identify the distributions and functions in a discrete choice model with unobserved heterogeneity, making minimal assumptions about the parametric structure of the systematic subutilities and no parametric structure about the distributions in the model. In particular, the theorem implies that, when some assumptions are satisfied, one can identify the functions $r(\cdot)$ and $h(\cdot)$ in Example 1, and the functions $h(\cdot)$, $t(\cdot)$, and $n(\cdot)$ in Example 2, without imposing any parametric structure in these functions.

4. NONPARAMETRIC ESTIMATION

Given N independent observations $\{y^i, s^i, z^i\}_{i=1}^N$ we can define the log-likelihood function:

$$L(V, \underline{F}, G) = \sum_{i=1}^{N} \log \int \left[\Pr(j|s^i, z^i; V(\cdot;\omega), \underline{F})\right]^{y^i_j} dG(\omega) \quad ,$$

where $\underline{F}= (F_1, ..., F_J)$. We then define our estimators, $\hat{V}$,$\hat{\underline{F}}$ and $\hat{G}$, for V^*, $\underline{F}^*$, and G^*, to be the functions and distributions that maximize $L(V, \underline{F}, G)$ over triples $(V, \underline{F}, G)$ that belong to a set $(W \times \Gamma_F \times \Gamma_G)$.

Let d_W, d_F, and d_G denote, respectively, metrics over the sets W, Γ_F, and Γ_G. Let $d : (W \times \Gamma_F \times \Gamma_G) \times (W \times \Gamma_F \times \Gamma_G) \to R_+$ denote the metric defined by

$$d[(V, \underline{F}, G), (V', \underline{F}', G')] = d_W(V, V') + d_F(\underline{F}, \underline{F}') + d_G(G, G').$$

Then, the consistency of the estimators can be established under the following assumptions:

ASSUMPTION 11: The set $(W \times \Gamma_F \times \Gamma_G)$ is compact with respect to the metric d.

ASSUMPTION 12: The metric d_W is such that convergence with respect to d_W implies pointwise convergence. And the metrics d_F and d_G are such that convergence with respect to d_F and d_G, respectively, imply a.e. convergence.

ASSUMPTION 13: $\forall$ $\underline{F} \in \Gamma_F$, F_j is continuous and strictly increasing, for all j=1,...,J.

ASSUMPTION 14: $\forall V \in W$, V is continuous and its value possesses an absolutely continuous distribution.

Using the arguments in Wald (1949), the following theorem is easily established:

THEOREM 2: Under Assumptions 1-14, $(\hat{V}, \hat{G}, \hat{\underline{F}})$ is a strongly consistent estimator of $(V^*, \underline{F}^*, G^*)$ with respect to the metric d.

The present the proof in the Appendix.

5. COMPUTATION

We next describe a method to calculate the estimators. The method is based on several facts that can be derived from the following optimization problem, which is the problem used to define the estimators:

$$\underset{(V,\underline{F},G)\in(W\times\Gamma_F\times\Gamma_G)}{Max} L(V,\underline{F},G) = \sum_{i=1}^{N} \log \int \left[\Pr(j|s^i,z^i;V(\cdot;\omega),\underline{F})\right]^{y_j^i} dG(\omega) \quad ,$$

First note that at any solution, $(\hat{V},\hat{\underline{F}},\hat{G})$, of this optimization problem, $\hat{G}$ will possess at most N points of support. We will denote the points of support of any G by $\omega_1,...,\omega_N$, and we will denote the masses that any G assigns to these points by $\pi_1,...,\pi_N$. Second, note that the value of the objective function depends on any function $V(\cdot,\omega)$ only through the values that $V(\cdot,\omega)$ attains at the finite number of observed vectors $\{(s^1,x_j^1,r_j)_{j=1,...,J},...,(s^N,x_j^N,r_j)_{j=1,...,J}\}$. Hence, since at a solution, $\hat{G}$ will possess at most N points of support, we will be considering the values of at most N different functions, $V(\cdot,\omega_c)$ $c=1,...,N$. I.e., we can consider for each j $(j=1,...,J)$ at most N different subutilities, $V(j,\cdot;\omega_c)$ $(c=1,...,N)$. For each i and c, we will denote the value of $V(j,s^i,x_j^i,r_j,\omega_c)$ by $V_{j,c}^i$. Next, we note that at a solution, the value of the objective function will depend on any $\underline{F}$ only through the values that $\underline{F}$ attains at the finite number of values $(V_{j,c}^i - V_{1,c}^i,...,V_{j,c}^i - V_{J,c}^i)$ $i=1,...,N,$ $j=1,...,J,$ $c=1,...,N$. We then let $F_{j,c}^i$ denote, for each j $(j=1,...,J)$, the value of a distribution function F_j at the vector $(V_{j,c}^i - V_{1,c}^i,...,V_{j,c}^i - V_{J,c}^i)$. It follows that a solution, $(\hat{V},\hat{\underline{F}},\hat{G})$, for the above maximization problem can be obtained by first solving the following finite dimensional optimization problem, and then interpolating between its solution:

$$\max_{\{V_{j,c}^i\},\{\pi_c\},\{F_{j,c}^i\}} \sum_{i=1}^{N} \log \sum_{c=1}^{N} \prod_{j=1}^{J} \left[F_{j,c}^i\right]^{y_j^i} \pi_c$$

subject to

$$(\{V_{j,c}^i\},\{\pi_c\},\{F_{j,c}^i\}) \in K$$

where K is a set of a finite number of restrictions on $\{V_{j,c}^i\},\{\pi_c\},\{F_{j,c}^i\}$. The restrictions characterize the behavior of sequences $\{V_{j,c}^i\},\{\pi_c\},\{F_{j,c}^i\}$ whose values correspond to functions V in W, probability measures G in Γ_G, and distribution functions F in Γ_F.

To see what is the nature of the restrictions determined by the set K, consider for example a binary choice model where $x_1 \in R_+$, $v(1,s,x_1,\omega) = r(s) + \omega\, x_1$, $v(2,s,x_2,\omega) = h(x_2,\omega)$, $r(0) = 0$, $h(0,\omega) = 0$ for all ω, $r(\cdot)$ is concave and increasing, and $h(\cdot,\cdot)$ is concave and decreasing. Then, the finite dimensional optimization problem takes the following form:

$$\max_{\{V^i_{j,c}\},\{\pi_c\},\{F^i_j\},\{T^i\},\{D^i_c\}} \sum_{i=1}^{N} \log \sum_{c=1}^{N} [F^i_c]^{y^i_1} \; [1 - F^i_c]^{(1-y^i_1)} \; \pi_c$$

subject to

(a) $F^i_c \leq F^k_d \quad if \quad r^i + x^i_1\, \omega_c - h^i_c \leq r^k + x^k_1\, \omega_d - h^k_d$, $\pi_c > 0$, and $\pi_d > 0$

(b) $0 \leq F^i_{j,c} \leq 1$,

(c) $r^i \leq r^k + T^k \cdot (s^i - s^k)$

(d) $h^i_c \leq h^k_c + D^k_c \cdot ((x^i_2, \omega_c) - (x^k_2, \omega_c)) \quad$ if $\quad \pi_c > 0$

(e) $T^k \geq 0, \quad r^{N+1} = 0, \quad s^{N+1} = 0,$

(f) $D^k_c \leq 0, \quad h^{N+1}_c = 0$, and $\quad x^{N+1}_2 = 0 \quad$ if $\quad \pi_c > 0$

for $i, k = 1, \ldots, N+1$; $c, d = 1, \ldots, N$; $j = 1, \ldots, J$.

Constraints (a) and (b) guarantee that the F^i_c values are those of an increasing function whose values are between 0 and 1. Constraint (c) guarantees that the r^i values correspond to those of a concave function. Constrains (d) guarantees that the h^i_c values correspond to those of a concave function, as well. Constraints (e) and (f) guarantee that the r^i and the h^i_c values correspond, respectively, to those of a monotone increasing and a monotone decreasing function, and that the

r^i and the h_c^i values correspond to functions satisfying $r(0) = 0$ and $h(0, \omega) = 0$ for all ω.

A solution to the original problem is obtained by interpolating the optimal values obtained from this optimization. (See Matzkin (1992,1993a,1995)) for more discussion of a similar optimization problem.)

To describe how to obtain a solution to this maximization problem, we let

$$\tilde{L}[(r^1, ..., r^{N+1}, T^1, ..., T^{N+1}), \{(h_c^1, ..., h_c^{N+1}, D_c^1, ..., D_c^{N+1})\}, (\pi_1, ..., \pi_N)]$$

denote the optimal value of the following maximization problem:

$$\max_{\{F_c^i\}} \sum_{i=1}^{N} \log \sum_{c=1}^{N} [F_c^i]^{y_1^i} \; [1 - F_c^i]^{(1-y_1^i)} \; \pi_c$$

subject to

(a) $F_c^i \leq F_d^k \quad if \quad r^i + x_1^i \, \omega_c - h_c^i \leq r^k + x_1^k \, \omega_d - h_d^k \quad , \pi_c > 0$, and $\pi_d > 0$

(b) $0 \leq F_c^i \leq 1$,

A solution to this latter problem can be obtained by using a random search over vectors $(F_c^1, ..., F_c^N)_{c=1,...,N}$ that satisfy the monotonicity constraint (a) and the boundary constraint (b).

Then, a solution to the full optimization problem can be obtained by using a random search over vectors $(r^1, ..., r^{N+1}, T^1, ..., T^{N+1})$, $\{(h_c^1, ..., h_c^{N+1}, D_c^1, ..., D_c^{N+1})\}$, and $(\pi_1, ..., \pi_N)$ that satisfy, respectively, constraints (c) and (e), constraints (d) and (f), and the following constraints:
$\pi_j \geq 0$ $(j = 1, ..., N)$ and $\sum_{j=1}^{N} \pi_j = 1$.

(See Matzkin (1993b) for a description of an algorithm that uses a random search over vectors satisfying constraints of the above type.)

In practice, one can find the optimal values the π_j's by first maximizing
$\tilde{L}[(r^1, ..., r^{N+1}, T^1, ..., T^{N+1}), \{(h_c^1, ..., h_c^{N+1}, D_c^1, ..., D_c^{N+1})\}, (\pi_1, ..., \pi_N)]$

subject to (c)-(f) and the constraint:
$\pi_1 = 1, \quad \pi_j = 0 \ (j = 2, ..., N)$.

Then,maximizing
$\tilde{L}[(r^1, ..., r^{N+1}, T^1, ..., T^{N+1}), \{(h_c^1, ..., h_c^{N+1}, D_c^1, ..., D_c^{N+1})\}, (\pi_1, ..., \pi_N)]$
subject to (c)-(f) and the constraint:
$\pi_1 \geq 0, \ \pi_2 \geq 0, \ \pi_j = 0 \ (j = 3, ..., N)$.

Next, maximizing
$\tilde{L}[(r^1, ..., r^{N+1}, T^1, ..., T^{N+1}), \{(h_c^1, ..., h_c^{N+1}, D_c^1, ..., D_c^{N+1})\}, (\pi_1, ..., \pi_N)]$
subject to (c)-(f) and the constraint:
$\pi_1 \geq 0, \ \pi_2 \geq 0, \ \pi_3 \geq 0, \ \pi_j = 0 \ (j = 4, ..., N)$.

and so on .

6. CONCLUSION

We have presented a method to estimate discrete choice models with unobserved heterogeneity. The method does not impose parametric assumptions either on the systematic subutility functions or on the distributions of the unobservable random vectors and the heterogeneity parameter. The estimators are computationally feasible and strongly consistent.

APPENDIX

PROOF OF THEOREM 1:

Assume w.l.o.g. that Assumption 1 holds for j=1. As it is well known, from the distribution of $\eta_j \equiv (\epsilon_1-\epsilon_j, ..., \epsilon_J-\epsilon_j)$ for some j, we can recover the distribution of η_k for all $k \neq j$ (see Thompson (1988). So, to establish the identification of $\underline{F}^*$, it is enough to determine the identification of F_1^*. To see that F_1^* is identified, let $(t_2, ..., t_J)$ be given. Let $(u_1, ..., u_J)$ be such that $(t_2, ..., t_J) = (u_1 - u_2 + \alpha_1 - \alpha_2, ..., u_1 - u_J + \alpha_1 - \alpha_J)$. Then, $\forall V \in W, G \in \Gamma_G$

$$\begin{aligned} F_1^*(t_2, ..., t_J) &= \int F_1^*(t_2, ..., t_J)\, dG(\omega) \\ &= \int F_1^*(u_1 - u_2 + \alpha_1 - \alpha_2, ..., u_1 - u_J + \alpha_1 - \alpha_J)\, dG(\omega) \\ &= \Pr(1|\bar{s}, \bar{x}_1, u_1, \bar{x}_2, u_2, ..., \bar{x}_J, u_J) \end{aligned}$$

where the last inequality follows from Assumption 5. It follows that F_1^* is identified, since if for some F_1 and $(t_2, ..., t_J)$, $F_1(t_2, ..., t_J) \neq F_1^*(t_2, ..., t_J)$, then $\forall V, V', G, G'$

$$p(1|\bar{s}, \bar{x}_1, u_1, \bar{x}_2, u_2, ..., \bar{x}_J, u_J; V, F, G) \neq p(1|\bar{s}, \bar{x}_1, u_1, \bar{x}_2, u_2, ..., \bar{x}_J, u_J; V', F^*, G')$$

Next, assume w.l.o.g. that the alternative, $\tilde{j}$, that satisfies Assumption 6 is $\tilde{j} = 2$, $\beta_2 = 0$, and the alternative, j^*, that satisfies Assumption 7 is $j^* = 1$. To show that $v^*(1, \cdot)$ and G^* are identified, we transform the polychotomous choice model into a binary choice model by letting $r_j \to -\infty$ for $j \geq 3$. Let $\eta \equiv \epsilon_2 - \epsilon_1$, and denote the marginal distribution of $\epsilon_2 - \epsilon_1$ by F_η^*. Since F^* is identified, we can assume that F_η^* is known. $\forall s, x_1, x_3, ..., x_J, r_1, r_2$,

$$\Pr(1|s, x_1, \tilde{x}_2, ..., x_J, r_1, r_2, ..., r_J) = \int F_\eta^*(v^*(1, s, x_1, \omega) + r_1 - r_2)\, dG^*(\omega).$$

Let $\gamma = v^*(1, s, x_1, \omega)$. Over any set where the values of s and x_1 are constant, v^* is increasing in ω. Let m^* denote the inverse of the function v^* over any set where the values of (s, x_1) stay constant. Since G^* is absolutely continuous, the distribution, S, of γ, conditional on (s, x_1), is also absolutely continuous. Let $g^*(\cdot)$

denotes the density of ω. Using the standard change of variables technique we get that $\forall u_1, u_2$

$$\begin{aligned}
\Pr(1|s, x_1, u_1, u_2) &= \int F_\eta^*(\gamma + u_1 - u_2)\, dG^*(\omega) \\
&= \int F_\eta^*(v^*(1, s, x_1, \omega) + u_1 - u_2)\, g^*(\omega)\, d\omega \\
&= \int F_\eta^*(\gamma + u_1 - u_2)\, g^*(m^*(1, s, x_1, \omega)) \left|\frac{\partial m^*(1, s, x_1, \omega)}{\partial \gamma}\right| d\gamma \\
&= \int F_\eta^*(\gamma + u_1 - u_2)\, dS^*(\gamma) \\
&= \int F_{\eta-\gamma}^*(u_1 - u_2)\, dS^*(\gamma)
\end{aligned}$$

By Teicher (1961), $S^*(\cdot)$ is identified. Let $f(s, x_1)$ be the marginal distribution of (s, x_1). $f(\cdot)$ is identified since it is just a marginal distribution of observable variables. Hence, since $S^*(\cdot)$ is the distribution of γ conditional on (s, x_1), we can identify the joint distribution, $f(\gamma, s, x_1)$, of (γ, s, x_1). Since ω is distributed independently of (s, x_1), it follows by Brown and Matzkin (1995) (see also Roehrig (1988)), that we can identify the function $v^*(1, \cdot)$ and the distribution of ω, from the joint distribution $f(\gamma, s, x_1)$, as long as

$$v(1, \cdot) \neq v'(1, \cdot) \Rightarrow \tfrac{\partial v(1, \cdot, \omega)}{\partial(s, x_1)} \neq \tfrac{\partial v'(1, \cdot, \omega')}{\partial(s, x_1)}$$

where ω and ω' are such that $y - v(1, s, x_1, \omega) = 0$ and $y - v'(1, s, x_1, \omega') = 0$.

In other words, identification will hold if for any $v(1, \cdot) \neq v'(1, \cdot)$ there exist some $(s, x_1, \omega, \omega')$ such that

$$v(1, s, x_1, \omega) = v'(1, s, x_1, \omega') \text{ and } \tfrac{\partial v(1, s, x_1, \omega)}{\partial(s, x_1)} \neq \tfrac{\partial v(1, s, x_1, \omega')}{\partial(s, x_1)}.$$

This will hold by Assumption 7. Hence, we can identify $G^*(\omega)$ and $v^*(1, s, x_1, \omega)$.

We now proceed to show that $v^*(j, \cdot)$ is identified for $j = 2, ..., J$. Since F^*, G^*, and $v^*(1, s, x_1, \omega)$ have been shown to be identified even when the functions $v^*(j, \cdot)$ ($j = 2, ..., J$.) are not, we can assume from this point on that F^*, G^*, and $v^*(1, s, x_1, \omega)$ are known.

If Assumption (8.i) is satisfied, then substituting each time $j = 1$ by $j = 3, 4, ..., J$ and following an analogous reasoning to that used to show that $v^*(1, s, x_1, \omega)$ is identified, it is easy to show that $v^*(j, s, x_j, \omega)$ $(j = 3, ..., J)$ is identified. Substituting $j = 2$ by $\hat{j}$ and substituting $j = 1$ by $j = 2$, the same reasoning yields that $v^*(2, s, x_2, \omega)$ is identified.

If Assumption (8.ii) is satisfied, then again $v^*(j, s, x_j, \omega)$ $(j = 3, ..., J)$ can be shown to be identified using previous arguments. To show that $v^*(2, s, x_2, \omega)$ is identified, we let

$\gamma = n_{2k}(s, x_2, x_k, \omega) \equiv v^*(2, s, x_2, \omega) - v^*(k, s, x_k, \omega)$.

Since n_{2k} is strictly increasing in ω, conditional on (s, x_2, x_k), we can use a similar argument as the one repeatedly used above (using Teicher (1961) result) to show that the joint distribution $f(\gamma, s, x_2, x_k)$ is identified. In particular, we can assume that for all t, $\Pr(\gamma \leq t | s, x_2, x_k)$ is known. Let $m_{2k}(s, x_2, x_k, t)$ denote the inverse function of n_{2k}, conditional on (s, x_2, x_k). Then, for all t,

$$\begin{aligned}\Pr(\gamma \leq t | s, x_2, x_k) &= \Pr(n_{2k}(s, x_2, x_k, \omega) \leq t | s, x_2, x_k) \\ &= \Pr(\omega \leq m_{2k}(s, x_2, x_k, t) | s, x_2, x_k) \\ &= G^*(m_{2k}(s, x_2, x_k, t)).\end{aligned}$$

Since G^* is strictly increasing, it follows that

$$m_{2k}(s, x_2, x_k, t)) = G^*(\Pr(\gamma \leq t | s, x_2, x_k)).$$

Hence, $m_{2k}(s, x_2, x_k, t)$ is identified, and therefore its inverse, $n_{2k}(s, x_2, x_k, \omega)$ is identified too. Since

$$n_{2k}(s, x_2, x_k, \omega) \equiv v^*(2, s, x_2, \omega) - v^*(k, s, x_k, \omega)$$

and $v^*(k, s, x_k, \omega)$ is identified,

$$v^*(2, s, x_2, \omega) = n_{2k}(s, x_2, x_k, \omega) + v^*(k, s, x_k, \omega)$$

is identified.

Finally, suppose that Assumption (8.iii) is satisfied. Then, we can use the identification of G^* and F^* to identify, as in the last argument, the difference functions

$$n_{kk+1}(s, x_k, x_{k+1}, \omega) \equiv v^*(k, s, x_k, \omega) - v^*(k+1, s, x_{k+1}, \omega) \quad k = 1, ..., M-1.$$

If $j_M = 1$,then using the identification of $v^*(M, \cdot)$ to identify $v^*(M-1, \cdot)$, using the identification of $v^*(M-1, \cdot)$ to identify $v^*(M-2, \cdot)$ using n_{M-2M-1}, and so on, one can identify $v^*(j_1, \cdot)$, $v^*(j_2, \cdot)$, ...,$v^*(j_{M-1}, \cdot)$.

If $j_M = 2$,then using the fact that $\forall j \ \forall(s, x_j, \omega)$

$$n_{j2}(s, x_j, \tilde{x}_2, \omega) \equiv v^*(j, s, x_j, \omega) - v^*(2, s, \tilde{x}_2, \omega) = v^*(j, s, x_j, \omega),$$

one can identify $v^*(j, s, x_j, \omega)$ since
$v^*(j, s, x_j, \omega) = n_{j2}(s, x_j, \tilde{x}_2, \omega)$.

Hence, $v^*(M-1, \cdot)$ is identified, and this can be used to identify, as before, the preceding members in the sequence. Finally, to identify $v^*(2, \cdot)$ we can use the fact that $n_{21}(\cdot)$ and $v^*(1, \cdot)$ are identified.

This completes the proof of Theorem 1.

PROOF OF THEOREM 2: We show the theorem by showing that the assumptions necessary to apply the result in Wald (1949) are satisfied. (See also Kiefer and Wolfowitz (1956).) For any $(V, G, \underline{F}) \in (W \times \Gamma_G \times \Gamma_F)$, define

$$f(y, z, V, G, \underline{F}) = \int \prod_{j=1}^{J} p(j|s, z; V(\cdot, \omega), \underline{F})^{y_j} \, dG(\omega)$$

and for any $\rho > 0$, define the function $f'(y, z, V, G, \underline{F}, \rho)$ by

$$f'(y, z, V, G, \underline{F}, \rho) = \sup_{d[(V,G,\underline{F}),(V',G',\underline{F}')]<\rho} f(y, z, V', G', \underline{F}').$$

We first show that

(a) $\forall (V, \underline{F}, G) \in (W \times \Gamma_F \times \Gamma_G)$ and $\forall \rho > 0$, $f'(y, z, V, G, \underline{F}, \rho)$ is a measurable function of (y, z).

Proof of (a): Clearly, it suffices to show that for all j,

$\sup_{d[(V,G,\underline{F}),(V',G',\underline{F}')]<\rho} \; p(j|s, z; V', F', G')$ is measurable in (s, z).

To show this, we note that since $(W \times \Gamma_F \times \Gamma_G)$ is a compact space, there exists a countable, dense subset of $(W \times \Gamma_F \times \Gamma_G)$. Denote this subset by $(\tilde{W} \times \tilde{\Gamma}_F \times \tilde{\Gamma}_G)$. By Assumption 12.

$\sup\{p(j|s, z; V', F', G') \mid d[(V, G, \underline{F}), (V', G', \underline{F}')] < \rho \ , \ (V', G', \underline{F}') \in (W \times \Gamma_F \times \Gamma_G) \ \} = \sup\{p(j|s, z; V', F', G') \mid d[(V, G, \underline{F}), (V', G', \underline{F}')] < \rho \ , \ (V', G', \underline{F}') \in (\tilde{W} \times \tilde{\Gamma}_F \times \tilde{\Gamma}_G)\}$

Since, suppose that the left hand side is bigger than the right hand side, then, there must exist $\delta > 0$ and $(V', G', \underline{F}') \in (W \times \Gamma_F \times \Gamma_G)$ such that $\forall (V'', G'', \underline{F}'') \in (\tilde{W} \times \tilde{\Gamma}_F \times \tilde{\Gamma}_G)$,

(i) $p(j|s, z; V', F', G') > \delta > p(j|s, z; V'', F'', G'')$.

But, $(\tilde{W} \times \tilde{\Gamma}_F \times \tilde{\Gamma}_G)$ is dense in $(W \times \Gamma_F \times \Gamma_G)$. Hence, there exists a sequence $\{(V_k, \underline{F}_k, G_k)\} \subset (\tilde{W} \times \tilde{\Gamma}_F \times \tilde{\Gamma}_G)\}$ such that $d[(V_k, \underline{F}_k, G_k), (V', \underline{F}', G')] \to 0$. Assumption 12 implies then that $p(j|s, z; V_k, \underline{F}_k, G_k) \to p(j|s, z; V', \underline{F}', G')$, which contradicts (i).

Hence, since

$\sup\{p(j|s, z; V', F', G') \mid d[(V, G, \underline{F}), (V', G', \underline{F}')] < \rho \ , \ (V', G', \underline{F}') \in (W \times \Gamma_F \times \Gamma_G) \ \} = \sup\{p(j|s, z; V', F', G') \mid d[(V, G, \underline{F}), (V', G', \underline{F}')] < \rho \ , \ (V', G', \underline{F}') \in (\tilde{W} \times \tilde{\Gamma}_F \times \tilde{\Gamma}_G)\}$

and $\forall (V', G', \underline{F}') \in (\tilde{W} \times \tilde{\Gamma}_F \times \tilde{\Gamma}_G), p(j|s, z; V', F', G')$ is measurable in (s, z), it follows that

$\sup\{p(j|s, z; V', F', G') \mid d[(V, G, \underline{F}), (V', G', \underline{F}')] < \rho \ , \ (V', G', \underline{F}') \in (W \times \Gamma_F \times \Gamma_G) \ \}$

is measurable in (s, z).

Next, we note that by Assumptions 12 and 14,

(b) $\forall \left\{(V_k, \underset{-}{F}_k, G_k)\right\}_{k=1}^{\infty}, (V, \underset{-}{F}, G)$ such that $\left\{(V_k, \underset{-}{F}_k, G_k)\right\}_{k=1}^{\infty} \subset (W \times \Gamma_F \times \Gamma_G), (V, \underset{-}{F}, G) \in (W \times \Gamma_F \times \Gamma_G)$, and $d\left[(V_k, \underset{-}{F}_k, G_k), (V, \underset{-}{F}, G)\right] \to 0$, one has that $\forall(y,z)$, except perhaps on a set of probability 0, $f(y, z, V_k, \underset{-}{F}_k, G_k) \to f(y, z, V, \underset{-}{F}, G)$.

Next, we note that by the definition of $f(y, z, V, G, \underset{-}{F})$,

(c) $E\left[f(y, z, V^*, G^*, \underset{-}{F}^*)\right] < \infty$,

and

(d) $\forall(V, \underset{-}{F}, G) \in (W \times \Gamma_F \times \Gamma_G)$ and $\rho > 0$ sufficiently small,

$E\left[f'(y, z, V^*, G^*, \underset{-}{F}^*, \rho)\right] < \infty$.

By (a)-(d) and Assumption 11, it follows from Wald (1949) that

$d\left[(\hat{V}, \hat{G}, \hat{\underset{-}{F}}), (V^*, G^*, \underset{-}{F}^*)\right] \to 0 \quad a.s.$

This completes the proof of Theorem 2.

REFERENCES

ALBRIGHT, R.L., S.R. LERMAN, and C.F. MANSKI (1977) "Report on the Development of an Estimation Program for the Multinomial Probit Model." Report for the Federal Highway Administration. Cambridge Systematics, Inc.: Cambridge, Massachusetts.

BROWN, D.J. and R.L. MATZKIN (1995) "Nonparametric Estimation of Simultaneous Equations, with an Application to Consumer Demand," mimeo, Northwestern University.

COSSLETT, S. R. (1983) "Distribution-free Maximum Likelihood Estimator of the Binary Choice Model, " Econometrica, 51(3), 765-782.

HECKMAN, J.J. (1974) "The Effect of Day Care Programs on Women's Work Effort," Journal of Political Economy.

-------------- (1981a) "Statistical Models for Discrete Panel Data," in C. Manski and D. McFadden (eds.), Structural Analysis of Discrete Data with Econometric Applications, M.I.T. Press.

-------------- (1981b) "Heterogeneity and State Dependence," in S. Rosen (ed.) Studies in Labor Markets, University of Chicago Press.

HECKMAN, J.J. and B. SINGER (1990a) "A Method for Minimizing the Impact of Distributional Assumptions n Economietric Models for Duration Data," Econometrica, pp. 271-320.

HECKMAN, J.J. and J. WALKER (1990a) "The Relationship Between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data," Econometrica, Vol. 58, No. 6, pp. 235-275.

-------------------------- (1990b) "Estimating Fecundability from Data on Waiting Times to First Conceptions," Journal of the American Statistical Association, Vol. 84, No. 408, pp. 958-965.

HECKMAN, J.J. and R. WILIS (1977) "A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women," Journal of Political Economy.

HAUSMAN, J.A. and D.A. WISE (1978) "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogenous Preferences," Econometrica, 46, 403-426.

HOROWITZ, J.L. (1992) "A Smooth Maximum Score Estimator for the Binary Choice Model," Econometrica, 60, 505-531.

ICHIMURA, H. and T.S. THOMPSON (1993) "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution," mimeo, University of Minnesota.

KIEFER, J. and J. WOLFOWITZ (1956) "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters," Annals of Mathematical Statistics, 27, 887-906.

KLEIN, R.W. and R.H. SPADY (1993) "An Efficient Semiparametric Estimator for Discrete Choice Models," Econometrica, 61, 387-422.

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

MATZKIN, R.L. (1991) "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," Econometrica, 59, 1315-1327.

____________(1992) "Nonparametric and Distribution-free Estimation of the Binary Choice and the Threshold Crossing Models," Econometrica, 60, 239-270.

____________(1993a) "Nonparametric Identification and Estimation of Polychotomous Choice Models," Journal of Econometrics, 58,137-168.

____________(1993b) "Computation and Operational Properties of Nonparametric Shape Restricted Estimators," mimeo, Northwestern University.

______________(1994) "Restrictions of Economic Theory in Nonparametric Methods," in Handbook of Econometrics, Vol. 4, by McFadden, D. and R. Engel (eds.).

TEICHER, H. (1961) "Identifiability of Mixtures," Annals of Mathematical Statistics.

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