

IDENTIFICATION OF SEMIPARAMETRIC

DISCRETE CHOICE MODELS

by

T. SCOTT THOMPSON

Discussion Paper No. 249, September 1989

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minn 55455

Keywords: semiparametric models, identification, discrete choice, random utility, isotonic functions, index models.

Headnote: The question of model identification is analyzed for the semiparametric random utility model of discrete choice. Attention is focused on settings where agents face a common choice between a set of $J+1$ alternatives, but where actual choices are only partially observed. Necessary conditions are derived for the setting where the only data on actual choices consists of a binary indicator for one of the alternatives. Sufficient conditions are developed in this setting for a linear in parameters specification of indirect utility. It is found that relative to the parametric case, only a mild continuity restriction on the distribution of regressors is needed in the semiparametric model. Under these circumstances all of the choice probabilities are identified, even though actual choices are only partially observed. It is shown that estimators that rely only on the index structure of the model require substantially stronger prior restrictions on the parameters for identification when the number of alternatives is large. Finally, results on the model with partial observability of choices are used to analyze the special case of full observability.

IDENTIFICATION OF SEMIPARAMETRIC DISCRETE CHOICE MODELS

by T. SCOTT THOMPSON¹

1. INTRODUCTION

The random utility model has become the canonical model in econometrics for situations where agents face a choice between a finite collection of discrete, unordered alternatives. In this model, for each alternative it is assumed that the indirect utility (conditional on choosing that alternative) is additively separable into a function of observables (the "regressors"), known up to a parametric family, and an unobserved disturbance (the "error" term). The actual choice is assumed to be the alternative for which the conditional indirect utility is at a maximum.

This paper examines the question of model identification under the assumption that the vector of errors is independent from observed determinants of indirect utility. We concentrate on the "semiparametric" version of this model in which no other restrictions are placed on the distribution of error terms. It is fruitful in this setting to consider the distribution function for the error terms to be a parameter of the model. Hence our identification results are informative about the error distribution in addition to the usual parameters of the observable part of indirect utility.

The primary focus of our investigation is on models with partial observability of choices. Specifically, the majority of the paper is devoted to the problem of identifiability of the model when the only data available

about actual choices is a binary indicator for whether or not one of the alternatives was chosen. The results for this case are extended in the last section to cover other settings, however. Full observability of choices is handled as a special case of the results there.

We also consider the question of identifiability through the index structure of the model. The random utility model of discrete choice is an "index" model because the conditional distribution of choices depends upon the observables through an index function. The particular relationship between the index function and the choice probabilities is determined by the error distribution function. We examine identification when one ignores this relationship and relies exclusively on the index property for identifying restrictions.

Relationship to Previous Studies.

Our analysis extends previous work in several directions. Of obvious importance is the relaxation of parametric assumptions. Often it is assumed that the collection of error terms is statistically independent from the observable determinants of indirect utility. In parametric versions of the model, which are the only versions that have been extensively used in applications, the vector of errors is also assumed to have a distribution that lies in a particular parametric family. Usually the Gaussian distributions or a generalized extreme value family of distributions are assumed, in which case the model respectively takes the familiar probit or nested logit structure. Extensive analysis of these models is contained in Daganzo (1983), Maddala (1983), McFadden (1974, 1976, 1981, 1982, 1984) and in numerous econometrics textbooks and monographs on discrete choice.

We also extend previous work on identification under partial observability. Poirier (1980) considered identification of bivariate normal probit models in an observational setting corresponding to our binary indicator for one of the possible choices. We extend this work by relaxing the normality assumption, by allowing for more general forms of partial observability, and by allowing for an arbitrary (finite) number of alternatives in the choice set.

More recently, a large number of estimators have been proposed for cases where the distribution of the errors is a priori unrestricted. We extend this literature primarily by allowing for multiple alternatives and partial observability. The semiparametric model has been considered by Cosslett (1983), Han (1984), Klein and Spady (1987), Manski (1975, 1985, 1988) and Matzkin (1988, 1990). Virtually all of this work analyzes only the binary choice setting. The exception is Manski (1975), who allowed for multiple alternatives, but who made the strong assumption that the error terms are independent and identically distributed across alternatives.

Our results on identification through index restrictions complement the growing body of literature on estimation of index models. The papers by Gallant (1987), Ichimura (1987), Ichimura and Lee (1990), Powell, Stock and Stoker (1987), and Stoker (1986, 1990) explore various approaches to estimation of general index models, of which discrete choice models form an important special case. Klein and Spady consider discrete choice models in particular, but their estimator only exploits the index structure of the model. Our results suggest that these estimators do not exploit all of the identifying restrictions on the parameters implicit in discrete choice models.

On the other hand, we maintain the assumption that the errors are independent from observables. Limited forms of interdependence between the error distribution and the regressors is allowed in much of the literature on parametric models, and in the papers by Manski for semiparametric models. Manski (1988) provides an extensive analysis of identification for the binary choice model under alternative assumptions about the joint distribution of observables and error terms.

We also avoid the complications introduced by individual variation in the set of feasible alternatives. While important, the added generality from this extension substantially complicates our analysis. We therefore leave it as a topic for future research.

Findings.

When the only data that is available about actual choices is a binary indicator for whether or not the first alternative is chosen it is necessary to normalize the vector of indirect utility functions against arbitrary permutations and rescalings of its components. Additional restrictions will be needed if the distribution of the regressors has a limited support.

Generic positive proofs of identification will often require a continuity condition on the distribution of the regressors when nothing is known about the error distribution. However, we show that in the linear in parameters model this "almost necessary" condition is the only new condition that needs to be added to the assumptions typically employed in parametric models in order to obtain a positive identification proof. Hence identification in semiparametric models requires only relatively weak (and verifiable) additional conditions beyond those needed for identification in

parametric models.

We find that if the parameters of the conditional choice probability function are identified in our partial information setting, then so are all of the conditional choice probabilities. This holds even though one cannot observe the actual choice except when the first alternative is chosen.

Our exploration of identification through index restrictions specifies the nature of the additional restrictions that are needed when one exploits only the index structure of the model. We find that the loss in identifying power from ignoring the full structure of the model is substantial whenever the number of alternatives is large.

Finally, we extend the results for observability of only one alternative to the case where one has full observability within a subset of the alternatives. It is found that if one can observe binary indicators for two or more of the possible choices then the relative scale of all of the indirect utility functions is identified. When the actual choice is fully observable then (as in parametric versions of the model) all of the parameters are identified up to a (common) positive scalar transformation.

Organization of the Paper.

Section 2 lays out a formal specification for the semiparametric random utility model with partial observability, develops notation, and provides some alternative interpretations for partial observability versions of the model. Section 3 analyzes various necessary conditions for identification. Section 4 discusses identification of the error distribution. Sufficient conditions for identification in the linear in parameters version of the model are covered in Section 5. Section 6 shows that under these conditions

all of the choice probabilities are identified. Section 7 shows that additional restrictions are needed on the parameters when one only exploits the index structure of the model for identifying restrictions. Section 8 extends the results from Section 5 to cases where additional information about actual choices is available. An Appendix provides proofs for the main results in the text.

2. THE SEMIPARAMETRIC RANDOM UTILITY MODEL OF DISCRETE CHOICE

Here we provide a formal specification for the semiparametric random utility model. We establish notation and define the formal statistical model to be analyzed in the remainder of the paper. We also consider some alternative interpretations of this statistical model provided by economic settings other than the traditional random utility model.

The Random Utility Model of Discrete Choice.

We assume that each member of a population of rational economic agents is faced with the problem of making a choice from a common finite set of $J+1$ alternatives labelled $0, \dots, J$. Each agent is (partially) characterized by a vector $x \in \mathbb{R}^L$ of observed (functions of) the characteristics of the agent and of the observed characteristics of the alternatives that are of relevance to the agent's decision problem. The observed choice c corresponds to the maximal element of an unobserved $(J+1) \times 1$ vector of (conditional indirect) utilities taking the form

$$(2.1) \quad \tilde{U} = \tilde{V}(x, \beta) + \tilde{\nu}$$

where \tilde{V} is a known function, β is an unknown parameter from a space B , and \tilde{v} is a vector of unobserved components of utility. That is, $c = j$ if and only if $\tilde{U}^{(j+1)} > \tilde{U}^{(i)}$ for each $i \leq J+1, i \neq j+1$.² It will also be convenient to define the random variables $y_j = 1(c = j), j = 0, \dots, J$.³

The distribution of characteristics, utilities and choices in the population is determined by regarding $\tilde{U}, x, \tilde{v}, c$ and y_0, \dots, y_J as random variables on an unspecified probability space with probability measure \mathbb{P} . We assume that x and \tilde{v} are independent under \mathbb{P} . The joint distribution of x and \tilde{v} determine the probabilities of all observable events. We assume that ties between the $\tilde{U}^{(j)}$ occur with zero probability under \mathbb{P} so that the choice c is uniquely determined by the preceding stochastic structure with probability one.

The Random Utility Model with Partial Observability.

Through most of this paper we will concentrate on the setting where the actual choice c is not observed. Instead the data consist of i.i.d. observations on the random variables (x, y_0) . Therefore, we assume that the only information about the actual choice that is available is an indicator for whether or not the first alternative was chosen.

There are several reasons for focusing on the partial observability setting. First, there exist settings in which choices are not fully observed. For example, economists are often able to determine whether or not a person has chosen to participate in the labor force but are not able to observe which of several alternative activities were chosen.

Second, the analysis of identification depends only on the

characteristics of the formal statistical model under analysis. Other settings of choice exist that do not fall into the utility maximization framework outlined above, but which generate a statistical model equivalent to that of the random utility model with partial observability. Hence the present results have implications for other models as well. Some of the alternative interpretations of our statistical model are mentioned below.

Third, results on identification for the partial observability model imply similar results for identification when the actual choice is fully observed. Since the labelling of the alternatives is arbitrary, identification of the full observability model can be accomplished by considering it to be a collection of $J+1$ partial observability models linked by restrictions on the parameters across the models. This approach is developed further in Section 8.

Choice behavior is completely determined by the ordinal properties of utility, so there is no hope of completely identifying a distribution for \bar{U} . Therefore it is convenient to transform the random utility model to a representation in terms of utility differences. Let

$$(2.2) \quad U = \begin{bmatrix} \bar{U}^{(1)} & - & \bar{U}^{(2)} \\ & \vdots & \\ \bar{U}^{(1)} & - & \bar{U}^{(J+1)} \end{bmatrix}, \quad V(x,b) = \begin{bmatrix} \bar{V}(x,b)^{(1)} & - & \bar{V}(x,b)^{(2)} \\ & \vdots & \\ \bar{V}(x,b)^{(1)} & - & \bar{V}(x,b)^{(J+1)} \end{bmatrix},$$

$$\text{and } \nu = \begin{bmatrix} \bar{\nu}^{(2)} & - & \bar{\nu}^{(1)} \\ & \vdots & \\ \bar{\nu}^{(J+1)} & - & \bar{\nu}^{(1)} \end{bmatrix}.$$

Then $U = V(x,b) - \nu$ gives the $J \times 1$ vector of utility differences relative to the first alternative. So $c = 0$ if and only if $U \geq 0$.⁴

Identification.

The most that one can learn from observations on y_0 and x is their joint distribution. We assume that the marginal distribution of x is uninformative about β or F_0 . So any information about these parameters implied by the data must be contained in the distribution of y_0 conditional on x . Since y_0 is a binary indicator, its conditional distribution is completely specified by the conditional choice probabilities

$$(2.3) \quad \begin{aligned} \phi_0(x) &= \Pr(U \geq 0 \mid x) = \Pr(v \leq V(x, \beta) \mid x) \\ &= F_0(V(x, \beta)) \text{ (a.s.)}. \end{aligned}$$

The study of identification will concentrate on whether or not (2.3) continues to hold when one substitutes alternative values for F_0 and/or β . Suppose that F_0 is known a priori to lie in some space \mathcal{F} of distribution functions. We adopt the following standard definition.

Definition: Identification. (F_0, β) is identified relative to (F, b) if and only if

$$(2.4) \quad \Pr(F(V(x, b)) - F_0(V(x, \beta))) < 1.$$

β is identified relative to b if and only if (F_0, β) is identified relative to (F, b) for each $F \in \mathcal{F}$. Similarly, F_0 is identified relative to F if and only if (F_0, β) is identified relative to (F, b) for each $b \in B$.

Hereafter when we say that F_0 or β are identified, we shall mean that they are identified relative to every other member of their respective parameter spaces \mathcal{F} and B . Notice that identification of F_0 does not imply

identification of β or conversely. Identification will generally depend upon the specifications for V and B and \mathcal{F} , upon the nature of F_0 and β , and upon the distribution of x .

Identification under our definition is necessary, but not sufficient for the existence of consistent estimators. Generally, one needs to establish that $F(V(x,b))$ is a smooth functional of (F,b) in some appropriate sense in order to establish consistency for an estimator of (F_0, β) .

Parameterizations of the Semiparametric RUM.

We shall assume that \mathcal{F} is the space of all cumulative distribution functions corresponding to probability measures on \mathbb{R}^J . For technical reasons it is often convenient to ensure that \mathcal{F} is compact. This can be achieved by redefining \mathcal{F} to be the space of all cumulative distribution functions corresponding to probability measures on the extended Euclidean space $\bar{\mathbb{R}}^J$.⁵ All of the results in this paper continue to hold on the larger parameter space provided that we maintain the assumption that F_0 corresponds to a proper probability distribution on \mathbb{R}^J . For economy of notation we shall not explicitly present the more general results. Several footnotes identify the specific modifications that are needed to accommodate this generalization.

Whenever possible we shall leave the specific forms of V and of the parameter space B unspecified. It should be understood that B is a topological space. In applications B will usually be a compact metric space. There is no need, in general, for B to admit a finite-dimensional parameterization, however. Therefore many of our results will apply to models in which no finite-dimensional parameterizations are imposed.

On the other hand, one can only obtain positive identification results

if one is able to place some structure on V and B . We will frequently adopt the following form, which is linear in the parameters β .

$$(C1) \quad V(x,b) = bx. \quad B \text{ is a set of } J \times L \text{ real matrices.}$$

This form is by far the most commonly used specification in applied work estimating random utility models. Usually restrictions are placed on the elements of B in the form of exclusion restrictions, equality constraints or inequality constraints. Since we must have $\beta \in B$, restrictions of this sort presume the availability of comparable prior information about β .

Example 1: McFadden (1974) proposed a model equivalent to the specification (C1) with B defined by

$$(2.5) \quad B = (I_j \otimes \theta', \theta \in \mathbb{R}^K, \|\theta\| = 1),$$

where $L = J \times K$.⁶ Here $b^{(j)} = [0 \mid \dots \mid \theta' \mid \dots \mid 0]$, where each zero is a $1 \times K$ row vector and where θ' is the j 'th $1 \times K$ subvector. This implies $V^{(j)}(x,b) = \theta' x_j$, where x_j is the j 'th $K \times 1$ subvector of x . This specification is appropriate when no purely personal characteristics of the decision makers are observed.

Equation (2.5) imposes $(J-1) \times K$ exclusion restrictions on each row of β , $(J-1) \times K$ equality constraints across the rows of β and a single scale normalization on β . We shall see that while certain restrictions on β are necessary, it is not necessary in general to provide such an explicit set of restrictions in order to ensure identification of β . Therefore we will usually avoid imposing explicit restrictions of this form on B .

Alternative Interpretations of the Model.

We noted above that alternative interpretations of the preceding statistical model are available. Rather than starting from the random utility model, one can take U , V , ν and y_0 as primitive elements of a general latent variable model of binary response under partial observability. Parametric versions of this model are analyzed by Poirier (1980) and McFadden (1984).

Under the alternative interpretation, it is assumed that there are J unobserved binary responses corresponding to the J events $\{ U^{(j)} \geq 0 \}$, but that we only observe a positive response if the J unobserved responses agree. In other words $y_0 = 1(U \geq 0)$.

For example, y_0 could be a binary decision made by a committee of J members for which each member has a veto. In this case, the j 'th member votes affirmatively if $U^{(j)} \geq 0$, but due to the veto an affirmative decision is only observed under unanimity.

Feinstein (1989) provides another example in a study of nuclear reactor safety monitoring. Here $U^{(1)} \geq 0$ whenever a safety violation occurs, $U^{(2)} \geq 0$ whenever an inspection occurs, and $y_0 = 1$ if and only if a violation is detected by the inspection. If the inspection process is perfect then $y_0 = 1$ if and only if $U \geq 0$.

Obviously there are many other problems that have this general form.

3. NECESSARY CONDITIONS FOR IDENTIFICATION

Here we develop a number of negative results concerning identification of β and F_0 . That is, we establish conditions under which there exist values of $(F,b) \in \mathcal{F} \times B$ that are observationally equivalent to (F_0,β) , but where $F \neq F_0$ or $b \neq \beta$. Since (F_0,β) cannot be identified unless none of these conditions hold, we implicitly establish necessary conditions for identification.

The results in this section all elaborate on the following fundamental proposition.⁷

Proposition 1: Suppose that there exists a function $H: \mathbb{R}^J \rightarrow \mathbb{R}^J$ and a $b \in B$ such that

$$(3.1) \quad V(x,b) = H(V(x,\beta)) \text{ (a.s.)}, \text{ and}$$

$$(3.2) \quad \nu \leq V(x,\beta) \iff H(\nu) \leq H(V(x,\beta)) \text{ (a.s.)}.$$

Then (F_0,β) is not identified relative to (F,b) , where F is the distribution function for $H(\nu)$.

Proposition 1 gives basic conditions under which (F_0,β) is not identified relative to (F,b) . So under the stated conditions β is not identified unless $b = \beta$. Similarly, unless $H(\nu) = \nu$ (a.s.) then $F \neq F_0$, and F_0 is not identified.

We next develop several results that imply necessary conditions for identification by establishing specific conditions under which Proposition 1 applies. These are presented in the form of several corollaries to Proposition 1. Several of the corollaries refer to the following class of

functions, which is the smallest class for which (3.2) must always hold.

Definition: Isotonic Function. A function $H: \mathbb{R}^J \rightarrow \mathbb{R}^J$ is isotonic if for each $u, v \in \mathbb{R}^J$,

$$(3.3) \quad u \leq v \iff H(u) \leq H(v).$$

The term 'isotonic' means 'order-preserving': Isotonic functions preserve the canonical partial ordering on \mathbb{R}^J . The composition $G \circ H$ of any two isotonic functions G and H is isotonic. Isotonic functions must be strictly increasing in each argument.⁸ Since the definition implies that $u = v \iff H(u) = H(v)$, isotonic functions are one-to-one. Hence if H is isotonic and onto then H^{-1} exists and is also isotonic. This holds in particular for all linear isotonic functions. We shall say that a $J \times J$ matrix is isotonic if and only if it defines a linear isotonic function.

Proposition 1, Corollary 1: Suppose that there exists an isotonic function H and a $b \in B$, $b \neq \beta$ such that

$$(3.4) \quad V(\xi, b) = H(V(\xi, \beta)) \text{ for each } \xi \in \mathbb{R}^L.$$

Then β is not identified. If also $\Pr(\nu \neq H(\nu)) > 0$ then F_0 is not identified.

Proposition 1, Corollary 2: If $V(x, b) = V(x, \beta)$ (a.s.) for some $b \in B$, $b \neq \beta$, then β is not identified.

Proposition 1, Corollary 3: Suppose there exists a $J \times J$ matrix $C \neq 0$ such that $CV(x, \beta) = \tilde{v}$ (a.s.). Suppose also that there exists an isotonic

function G , and a $b \in B$, $b \neq \beta$, such that (3.4) holds for $H = G + C$. Then β is not identified. If also $\Pr(\nu \neq G(\nu) + \bar{v}) > 0$ then F_0 is not identified.

Proposition 1, Corollary 4: Suppose that

(3.5) the support of x is the finite set $\{x_1, \dots, x_n\}$,

(3.6) $V(x_s, \cdot)$ is continuous at β , $s = 1, \dots, n$, and

(3.7) $V^{(j)}(x_s, \beta) \neq V^{(j)}(x_t, \beta)$, $s, t \leq n$, $s \neq t$, $j \leq J$.

Then β is not identified. If also

(3.8) $\Pr(V(x, \beta) = V(x, b)) < 1$ for every b in a neighborhood of β ,
 $b \neq \beta$, or

(3.9) $\sum_{s=1}^n \Pr(\nu^{(j)} = V^{(j)}(x_s, \beta)) < 1$ for some $j \leq J$,

then F_0 is not identified.

Proposition 1, Corollary 5: Suppose there exists a $J \times 1$ vector c with exactly one strictly positive element $c^{(i)}$ such that $c'U \geq 0$ (a.s.). Without loss of generality let $c^{(i)} = 1$. Let H be the $J \times J$ matrix that coincides with the identity except for its i 'th row, which is given by $H_{ii} = 0$ and $H_{ij} = -c^{(j)}$ for $j \neq i$. If there exists a $b \in B$, $b \neq \beta$, such that (3.4) holds for this H then β is not identified. If $\Pr(c'\nu = 0) < 1$ then F_0 is not identified either.

Corollary 1 shows that restrictions on V or B are needed to normalize against isotonic transformations of $V(\cdot, \beta)$. That is, the following condition is necessary for identification of β .

(C2) For each $b \in B$, if there exists an isotonic H such that
 $V(\xi, b) = H(V(\xi, \beta))$ for every $\xi \in \mathbb{R}^L$ then $b = \beta$.

The practical significance of this result depends directly on the nature of the class of isotonic functions. For example, the identity function is isotonic. Then Corollary 1 establishes the trivial result that β is not identified if the functions $V(\cdot, \beta)$ and $V(\cdot, b)$ coincide for some $b \neq \beta$. This indicates that a more parsimonious parameterization of V exists. We shall not consider it further.

The class of isotonic functions is not trivial. All functions of the form $H(v) = v + \bar{v}$ (\bar{v} constant) are included. So are functions satisfying $H(v)^{(j)} = h_j(v^{(j)})$ for some collection h_j , $j \leq J$, of strictly increasing one-to-one and onto functions. In particular, linear scaling functions of the form $H(v) = Dv$, where D is a positive-definite diagonal matrix, are included in the class. The permutation functions $H(v) = Pv$ for some $J \times J$ permutation matrix P are isotonic. More complicated functions can be constructed from these examples by forming the composition of two or more isotonic functions.

From these examples Corollary 1 establishes the well-known fact that scale and location normalizations must be placed on the indirect utility function to ensure identification of β . A location normalization is required because a shift in the centering of the F_0 distribution cannot be distinguished from a change in a constant term in V . Hence the model specification does not provide a natural choice for the origin of \mathbb{R}^J . Similarly, deformation (or rescaling) of any one of the coordinate axes cannot be distinguished from a monotone transformation of the corresponding component of V . So scale and/or shape restrictions are needed on each

component of V .

Corollary 1 also demonstrates the need to place restrictions that guard against permutations of the coordinates of V . This restriction is not needed when actual choices are observable. In the partial observability setting considered here, however, permitting permutations of the coordinates leads to a failure of identification. Since the distribution of observables is presumed to embody information on actual choices only through the binary variable y_0 , one cannot distinguish between the remaining alternatives when the first alternative is not chosen. Hence any arbitrary relabeling of the other choices, corresponding to a permutation of the coordinates of V , is fully consistent with the partial observability specification of the random utility model.

The need for scale and permutation normalizations arises from the fact that choices are only partially observed rather than from its semiparametric specification. Scale and permutation normalizations are also necessary in parametric versions of the random utility model with partial observability. See, for example, Poirier (1980), who considered the case where F_0 is assumed to be bivariate normal.⁹

The failure of identification in Corollary 1 occurs regardless of the stochastic structure of the observables.¹⁰ The conditions of Proposition 1, however, make clear that identification can also fail if the support of $V(x, \beta)$ is too small. Corollaries 2 - 5 identify specific conditions under which this occurs. In particular, Corollaries 2 - 4 identify cases in which $V(x, b)$ almost surely equals an isotonic transformation of $V(x, \beta)$. The conditions of these Corollaries may hold even though (3.4) is not satisfied for any isotonic function H .

Corollary 2 extends Corollary 1 to cover cases where the functions $V(\cdot, b)$ and $V(\cdot, \beta)$ differ but nevertheless coincide on the support of x . In this case the variation in x is insufficient to identify β from other members of B . This is equivalent to the failure of identification associated with extreme multicollinearity in the linear regression model. As in the case of multicollinearity, the only solution is to impose further restrictions on B .

Corollary 3 shows that β may not be identified if the support of $V(x, \beta)$ is contained in a proper affine subspace of \mathbb{R}^J .¹¹ Identification fails because $V(x, b)$ almost surely equals an isotonic transformation of $V(x, \beta)$. Since the function $G + C$ need not be isotonic, the Corollary shows that in this case restrictions on β (and B) may be needed beyond the normalizations against isotonic transformations identified in Corollary 1.

On the other hand, if $CV(\cdot, \beta)$ is identically zero then $H(V(\cdot, \beta))$ coincides with $G(V(\cdot, \beta))$. This case is covered by Corollary 1, since G is isotonic. So Corollary 3, like Corollary 2, contributes something new only when the degeneracy in the support of $V(x, \beta)$ is caused by insufficient variation in x .¹² As in the cases considered by Corollary 2, the only solution is to obtain additional restrictions on B .

The literature on parametric random utility models typically makes assumptions that ensure that the support of $V(x, \beta)$ is not contained in any proper affine subspace of \mathbb{R}^J . This has obscured the fact that these models are also susceptible to failures of identification of the sort described by Corollary 3. Consider the following example, which is equivalent to a version of the model considered by Poirier (1980).

Example 2: Let $J = L = 2$. Restrict \mathcal{F} to be the space of all bivariate normal distribution functions (with arbitrary mean vectors and covariance matrices). Suppose that (C1) holds and that B is the set of all 2×2 matrices of the form

$$(3.10) \quad b = \begin{bmatrix} 1 & b_{12} \\ b_{21} & -1 \end{bmatrix}, \quad b_{12} \geq 0, \quad b_{21} \leq 0. \quad ^{13}$$

Let $\beta = I$.¹⁴ Suppose also that $x_2 = -x_1$ (a.s.) and that x_1 has an everywhere positive Lebesgue density. Let

$$(3.11) \quad G = \begin{bmatrix} 0 & .5 \\ .5 & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Let $H = G + C$. Then

$$(3.12) \quad H = \begin{bmatrix} 1 & -.5 \\ -.5 & 1 \end{bmatrix}, \quad \text{and} \quad \tilde{b} = H\beta = \begin{bmatrix} 1 & .5 \\ -.5 & 1 \end{bmatrix} \in B.$$

Letting $F = F_0 \circ G^{-1}$, one can verify directly that $F(\tilde{b}x) = F_0(\beta x)$ (a.s.), so (F_0, β) is not identified relative to (F, \tilde{b}) . Furthermore $F \in \mathcal{F}$, since G^{-1} is linear and since \mathcal{F} is invariant under linear transformations of the underlying bivariate normal random vectors. So neither β nor F_0 are identified.

The failure of identification in the preceding example is partially due to the degenerate distribution of x . Even so, B has been sufficiently restricted to ensure that there is no $b \in B$, $b \neq \beta$, for which $V(x, b) = V(x, \beta)$ (a.s.). So β would be identified in this example if U were directly observed. Therefore the example shows that when x has a degenerate

distribution, identification of β requires more restrictions on B than are required in the linear regression model.

Corollary 4 shows that identification will often fail if x has finite support. Under Conditions (3.5) - (3.7) it is possible to find a $b \in B$ close to β such that $V(x,b)$ almost surely equals a nonlinear rescaling of $V(x,\beta)$. Since nonlinear rescalings are isotonic, it is then easy to verify the conditions of Proposition 1. The failure of identification analyzed by Corollary 4 is local in the sense that β may still be identified relative to values of b that are sufficiently far from β .

Corollary 4 can be regarded as an extension of previous results for parametric versions of the random utility model. Poirier (1980) noted that in parametric versions of the model a necessary condition for identification is for the support of x to have a cardinality at least as great as the number of free parameters in the model.¹⁵ No finite-dimensional parameterization exists for the semiparametric random utility model, however. The implicit number of (scalar) free parameters in this model is infinite. Therefore Poirier's observation suggests that identification might fail in the semiparametric model if the support of x has finite cardinality.¹⁶ The Corollary shows that this is indeed the case for certain smooth parameterizations.

Corollary 5 shows that if the support of U is bounded in certain directions then there may exist linear functions H that are not isotonic but which nevertheless satisfy the conditions of Proposition 1.¹⁷ This problem does not arise in most parametric models since these models typically assume that the support of ν (hence of U) covers \mathbb{R}^J .

Notice that $c'U \geq 0$ if and only if $c'V(x,\beta) \geq c'\nu$. Since one can

rarely say anything about the support of ν a priori, the only way to guarantee that the conditions of the Corollary do not hold is to establish that $\Pr(c'V(x,\beta) \geq \gamma) < 1$ for each vector c of the form specified in the Corollary and for every real number γ . This is equivalent to establishing that the support of $V(x,\beta)$ is not contained in certain halfspaces of \mathbb{R}^J that intersect the non-negative orthant in non-trivial ways. Alternatively, one can place further restrictions on B to ensure that (3.4) does not hold for any H of the form specified in the Corollary.

Identification fails in the manner of Corollary 5 because the underlying structure of preferences is degenerate. To see this, consider the case where the conditions of the Corollary are satisfied for a vector c with exactly one element equal to 1, one element equal to -1, and the remaining elements all zero. For example, suppose that $c^{(1)} = -c^{(2)} = 1$. Then $c'U \geq 0$ (a.s.) if and only if

$$(3.13) \quad U^{(1)} - U^{(2)} - \bar{U}^{(3)} - \bar{U}^{(2)} \geq 0 \text{ (a.s.)}.$$

That is, the third alternative is preferred to the second with probability one. In this case the first alternative is preferred to the second alternative if and only if it is also preferred to the third alternative, so an equivalent model is obtained when $\bar{U}^{(2)}$ is replaced by $\bar{U}^{(3)}$. The transformation H has exactly this effect: $H^{(1)}U - H^{(2)}U - U^{(2)} - \bar{U}^{(1)} - \bar{U}^{(3)}$. Clearly strong prior information about β will be needed to sort out the choice behavior in this case.

An "Almost Necessary" Condition.

Unlike Corollary 1, the conditions in Corollaries 2 - 5 may or not hold depending upon the particular stochastic structure of the random utility model. In each case identification may still be achieved if sufficient prior information is available about β to allow certain restrictions to be imposed on B. Unfortunately, the particular restrictions that are needed will depend upon the specific nature of the parameterization of V and B, on the true value β and on the distribution of x. Hence these Corollaries do not immediately suggest additional generic necessary conditions for identification.

The common feature of the conditions in Corollaries 3 - 5, however, is a restricted support for $V(x, \beta)$. The conditions of these Corollaries all fail if the following holds.

$$(C3) \quad \Pr(V(x, \beta) \in E) > 0 \text{ for every open set } E \subset \mathbb{R}^J.$$

This continuity condition ensures that the support of $V(x, \beta)$ coincides with \mathbb{R}^J .

Condition (C3) certainly is not necessary for identification. It is "almost necessary," however, in the sense that weaker conditions will require restrictions on V, β and x that do not hold in general, but which must be verified on a case-by-case basis. So positive, generic proofs of identification can generally be obtained only under conditions that imply (C3). The discussion in Section 4 will provide some additional reasons for imposing (C3).

Virtually all previous work on estimation of semiparametric models of discrete choice assumes Condition (C3) directly, or else assumes conditions that imply (C3).

Linear Transformations and Identification.

When does a function H satisfy the conditions of Proposition 1? In particular, for which H does (3.2) hold? The Corollaries to Proposition 1 provide partial answers to these questions by providing specific examples. Here we address a closely related question: What properties must H possess if it satisfies equation (3.2)?

We have no general answer to this question. Instead, the following propositions provide a specific answer for the case where H is restricted to be linear.

Proposition 2: Suppose that $H: \mathbb{R}^J \rightarrow \mathbb{R}^J$ is linear and that $F_0 = F \circ H$ for some $F \in \mathcal{F}$.¹⁸ Then H is isotonic and $F = F_0 \circ H^{-1}$.

To see the significance of this Proposition, suppose that $V(x, b) = HV(x, \beta)$ (a.s.) for some linear transformation H . If also (F_0, β) is not identified relative to (F, b) then we must have $F(HV(x, \beta)) = F_0(V(x, \beta))$ (a.s.). Then if the support of $V(x, \beta)$ is large enough (if Condition (C3) holds, for example) then we must have $F_0 = F \circ H$. Proposition 2 establishes that the last condition holds only if H is isotonic, in which case $F = F_0 \circ H^{-1}$ provides an explicit expression for F . Therefore, under these conditions the only linear transformations that can satisfy (3.2) are isotonic. It follows that if Conditions (C2) and (C3) hold then $V(x, \beta)$ is identified relative to every linear transformation of itself.

On the other hand, a more complete characterization of isotonic functions will be needed in order to implement the normalizations specified

by Condition (C2). We have already identified a number of specific examples of isotonic transformations. It is not known if others exist. It turns out, however, that these examples exhaust all possibilities when attention is restricted to linear isotonic functions. Proposition 3 formalizes this result.

Proposition 3: A linear transformation H is isotonic if and only if there exist square matrices P and D such that $H = PD$, where P is a permutation matrix and D is a positive-definite diagonal matrix. Equivalently, H is isotonic if and only if there is exactly one strictly positive element in each row and in each column of H , with the remaining elements all equal to zero.

Proposition 3 shows that scale and permutation normalizations are sufficient to exclude isotonic transformations of $V(x, \beta)$ (other than the identity) whenever V is linear in the parameters. No other normalizations on β will be necessary in many applications.

4. IDENTIFICATION OF THE ERROR DISTRIBUTION

Distribution functions satisfy certain monotonicity, additivity and continuity properties. Given these restrictions, it may be possible to completely determine F_0 from its values on some subset of \mathbb{R}^J . For example, F_0 is completely determined by the values it takes on its support. So F_0 may be identified if $V(x, \beta)$ is identified and has a rich enough distribution. The following provides the first positive identification result along these lines.

Proposition 4: Let V_0 be the support of $V(x, \beta)$, let $\text{int}(V_0)$ denote its interior and let S_0 be the support of F_0 . If β is identified and $S_0 \subset \text{int}(V_0)$ then F_0 is identified.

If β is identified but it is not the case that $S_0 \subset \text{int}(V_0)$ then F_0 generally will not be identified. For example, if $V(x, \beta)$ has a degenerate distribution so that $\text{int}(V_0)$ is empty, but $S_0 = \mathbb{R}^J$, then there will be open subsets of \mathbb{R}^J on which F_0 can be bounded in various ways, but not enough to identify an individual distribution function. The proof of the Proposition reveals that F_0 always is identified relative to those $F \in \mathcal{F}$ that do not coincide with F_0 on $\text{int}(V_0)$, however. We are unaware of more general criteria on S_0 and V_0 that ensure identification of F_0 relative to every other element of \mathcal{F} .

Practical use of the Proposition requires prior knowledge of the supports of F_0 and $V(x, \beta)$. Typically, only weak a priori restrictions are available on the support of F_0 . Often all one knows is that F_0 corresponds

to a proper probability distribution. Then conditions are needed that guarantee that the support of $V(x, \beta)$ is all of \mathbb{R}^J . Therefore Condition (C3) will usually be needed to establish a generic, positive proof of identification of F_0 .

5. SUFFICIENT CONDITIONS FOR IDENTIFICATION: THE LINEAR IN PARAMETERS MODEL

This section develops a set of specific conditions under which β and F_0 are identified. A complete set of sufficient conditions requires more specific structure on the parameterization of the semiparametric random utility model than we have assumed so far. Therefore attention will be restricted to the linear in parameters specification (C1). Proposition 5 presents the main result.

Proposition 5: Suppose that in addition to Condition (C1) the following conditions hold:

- (C4) The support of x is not contained in any proper affine subspace of \mathbb{R}^L .¹⁹
- (C5) $\beta x = \beta_1 x_1 + \beta_2 x_2$ where β_1 is a nonsingular $J \times J$ submatrix of β and $\Pr(x_1 \in E \mid x_2) > 0$ almost surely for every open set $E \subset \mathbb{R}^J$.

If also $F(V(x, b)) = F_0(V(x, \beta))$ (a.s.) for some $(F, b) \in \mathcal{F} \times B$ then $b = H\beta$ and $F = F_0 \circ H^{-1}$ for some isotonic matrix H .

Proposition 5, Corollary: Suppose Conditions (C1), (C4) and (C5) hold.
Suppose also that

(C6) For each $b \in B$, if $b = PD\beta$ for some permutation matrix P and
some positive-definite diagonal matrix D then $P = D = I$.

Then (F_0, β) is identified.

The remainder of this section provides some interpretation for these
sufficient conditions.

It is well-known that linear independence of the components of x is
necessary for identification in parametric models satisfying Condition (C1).²⁰
Condition (C4) imposes the stronger requirement that these components are
affinely independent as well. This means that x cannot include a constant
term, either explicitly or through some nontrivial linear combination. In
particular, x cannot include any characteristics that are purely alternative
specific (the price of an alternative, for example), since there is no
variation in these characteristics across individual decision-makers.

Closer examination reveals that the exclusion of constants from x in
the semiparametric random utility model is equivalent to the location
normalization imposed on F_0 in virtually all parametric implementations of
the model. In parametric models the location normalization is provided by
some centering restriction on F_0 . Typically the univariate marginal
distributions of F_0 are required to be symmetric about zero. In the
semiparametric setting F_0 is unrestricted. Given any location measures (such
as the medians) for the random variables $\nu^{(j)}$, the semiparametric model can
be regarded as having an implicit constant term with coefficient in $V^{(j)}$
provided by the location measure for $\nu^{(j)}$. Condition (C4) specifies that the

components of x cannot be collinear with this implicit constant term.

Constant terms can be reintroduced to the semiparametric model if a location normalization is imposed on F_0 . Alternatively, given Condition (C4), one can use location measures on the univariate marginal distributions of F_0 to provide estimates of purely alternative specific effects on the choice probabilities.

When the marginal distributions of F_0 are symmetric, all reasonable location measures provide identical estimates of these effects. In the semiparametric setting, however, one does not know that F_0 is symmetric, so different location measures may provide different measures of alternative-specific effects. We choose to leave F_0 free and impose the location normalization through Condition (C4) in order to avoid choosing among these different measures.

The other normalizations are provided by Condition (C6), which excludes nontrivial isotonic transformations of β from the parameter space. Given the linear specification, Proposition 3 shows that only scale and permutation normalizations on the rows of β are needed to satisfy this condition.

The need for these restrictions has already been established but their implementation has not. Consider Example 1. The definition of B in (2.5) normalizes against permutations by imposing $(J-1) \times K$ exclusion restrictions on each row of b . Scale normalizations are provided by the normalization $\|\theta\| = 1$ together with the $(J-1) \times K$ equality constraints across the rows of b .

In general, scale normalizations on B can be imposed by requiring $\|b^{(j)}\| = 1$ for each j . While exclusion restrictions can certainly be used to normalize against permutations, equality constraints are not necessary.

Certain inequality constraints across the rows of β will work equally well. For example, if one knows a priori that $\beta_{11} < \beta_{21} < \dots < \beta_{J1}$ then the restrictions $b_{11} \leq b_{21} \leq \dots \leq b_{J1}$ can be used to normalize against permutations.

Often, however, economic theory provides no a priori justification for imposing constraints that normalize against permutations.²¹ Fortunately, it is rarely necessary to provide an explicit a priori normalization. Poirier (1980) notes that while permutation normalizations are needed for global identification, they are unnecessary for local identification.²² So there is no need to impose these normalizations explicitly when employing local search algorithms to compute estimates of β . Furthermore, inferential exercises involving β sometimes are invariant to permutations of its rows. Hence the permutation normalization will often be ignored in practice.

We have seen that Conditions (C1), (C4) and (C6) are each equivalent to conditions routinely assumed for parametric implementations of the random utility model. Therefore the new restrictions needed for Proposition 5 in order to establish identification for the semiparametric model are all contained in Condition (C5). This condition imposes additional structure jointly on β and on the distribution of x . Conditions (C4) and (C5) together imply (C3). Hence (C5) is used in part to avoid some of the identification difficulties identified in the Corollaries to Proposition 1.

Condition (C5) can be analyzed in several parts. First, β must include a $J \times J$ nonsingular submatrix β_1 , which implies that β has full row rank. Second, there must be a decomposition of x into subvectors x_1 and x_2 such that the support of x_1 covers \mathbb{R}^J conditional on x_2 . Third, the elements of x that comprise x_1 must correspond to the columns of β that comprise β_1 .

Taken as a whole, these requirements are somewhat restrictive. Clearly any discrete components of x must be included in x_2 . While the condition does not require that x_1 (conditionally) possesses a Lebesgue density, it is not sufficient for each component of x_1 to be absolutely continuous. In particular, the condition requires that there not exist any function g such that $g(x) = 0$ almost surely, unless g can be written as a function of x_2 alone. Hence no elements of x_1 can be deterministically related to other components of x through interaction or polynomial expansion terms. So the condition restricts somewhat the ability of the linear in parameters specification to include specifications that are nonlinear in the data.

The condition does not require a priori knowledge of which submatrix of β comprises β_1 . For example, if x possesses an everywhere positive Lebesgue density on \mathbb{R}^L then (C5) will be satisfied for one of the $J \times 1$ subvectors of x provided that β has full row rank. Often, however, not all subvectors of x will satisfy the restrictions needed on x_1 . Then the condition implies rank restrictions on certain submatrices of β . Since the condition places no restrictions on values of $b \in B$ other than the true value β , there is no need to impose these rank restrictions in the definition of the parameter space B .

Condition (C5) is not necessary for identification. In the case $J = 1$, for example, there exist cases where F_0 and β are identified despite the fact that no component of x is continuous conditional on the other components. We are not aware of any generally applicable alternative condition, however. It does not appear that the condition can be substantially weakened except in special cases involving stronger prior information about F_0 or β .

On the other hand, Condition (C5) corresponds closely to other conditions commonly imposed in the literature on semiparametric estimation of

nonlinear regression models. For example, if $J = 1$ then Condition (C5) is equivalent to Manski's (1988b) Condition (X3) for identification of binary response models.

6. IDENTIFICATION VIA INDEX RESTRICTIONS

A conditional regression model is said to have the "index" property if the distribution of y conditional on x depends on x only through some parametric "index" function $V(x, \beta)$. The semiparametric random utility model with partial observability is embedded in this class of models since the conditional choice probabilities for the first alternative are given by $\phi_0(x) = F_0(V(x, \beta))$. Notice, however, that no structure on F_0 is assumed for general index models, while the random utility model requires that F_0 is a distribution function.

This section considers the loss of identifying information (if any) that occurs when one ignores the fact that F_0 is a distribution function. The motivation for this exercise is provided by the growing collection of "index" estimators that have been proposed for the semiparametric random utility model. These estimators have the property that they are consistent for β (under conditions resembling those of Proposition 5) even when F_0 is not a distribution function. Hence they implicitly ignore the identifying power available in the restriction $F_0 \in \mathcal{F}$.

Examples of "index" estimators for the case $J = 1$ include Ichimura (1987), Powell, Stock and Stoker (1987), Stoker (1987, 1990), and Klein and Spady (1987). The last applies only in the special case where y is a

Bernoulli random variable conditional on x . Ichimura and Lee (1990) proposed an estimator that applies to index models when $J \geq 1$. We shall generically call these estimators "index estimators."

Comparison of the identification results in Section 5 to those in the preceding papers shows that relying exclusively on the index property for identification results in a substantial loss of identifying power. This is because, given any invertible function $H: \mathbb{R}^J \rightarrow \mathbb{R}^J$, one can always write $\phi_0(x) = F(H(V(x, \beta)))$, where $F = F_0 \circ H^{-1}$. We have seen that this F is a distribution function if and only if H is isotonic. If one ignores the restriction $F_0 \in \mathcal{F}$, however, then $V(x, \beta)$ is clearly not identified relative to any invertible transformation of itself.

Significantly stronger prior information is needed to identify $V(x, \beta)$ relative to arbitrary invertible transformations than is needed to identify it only relative to isotonic transformations. To see this, consider the linear in parameters model. Suppose that β is not identified relative to $b = H\beta$ for some nonsingular matrix H . If $b \in B$ and β has full row rank then β is not identified unless B is restricted sufficiently to ensure that $H = I$. So one can quantify the loss in identifying power that results from ignoring $F_0 \in \mathcal{F}$ by comparing the additional restrictions needed on H to ensure $H = I$ when H is known to be isotonic with those needed when H is known only to be nonsingular.

Now a matrix H is nonsingular if and only if

$$(6.1) \quad H = \text{PLSDU},$$

where P is a $J \times J$ permutation matrix, L and U are respectively lower and upper triangular $J \times J$ matrices with ones along the main diagonal, D is a $J \times J$ positive-definite diagonal matrix, and S is a $J \times J$ diagonal "sign" matrix,

whose diagonal elements each equal plus or minus one. Proposition 3 establishes that H is isotonic if and only if

$$(6.2) \quad L - S - U = I.$$

But it is easy to establish that $H = I$ if and only if (6.2) holds and also

$$(6.3) \quad P - D = I.$$

So (6.2) represents the additional restrictions needed on each element of B in order to ensure that β is identified when one ignores that $F_0 \in \mathcal{F}$.

If the restriction $F_0 \in \mathcal{F}$ is ignored then one needs J^2 a priori restrictions on β (corresponding to the free parameters of L , S and U) in order to completely identify β .²³ Of these, J will be sign restrictions. So when $J = 1$ the cost from ignoring $F_0 \in \mathcal{F}$ is a single sign restriction on β .

For $J \geq 2$, equality restrictions will be needed on β beyond the innocuous restrictions needed to normalize against scale and permutation transformations. In these cases one needs substantially stronger prior information in order to successfully identify β using index restrictions than is needed to identify β using the full structure of the random utility model. Since the off-diagonal elements of L and U can vary continuously, failure to impose these additional constraints will lead to a failure of local and global identification.

It is true that the index estimators are robust to deviations from the random utility model in the form of heteroskedastic disturbances for ν . In order for the index property to hold the distribution of ν conditional on x can depend on x only through $V(x, \beta)$, however. It is questionable whether this limited form of robustness compensates for the possibility of misspecification that arises when unnecessarily strong, and perhaps incorrect, restrictions are placed on β .

7. IDENTIFICATION OF THE CHOICE PROBABILITIES

So far we have concentrated on identification of F_0 and β . In this section we establish that identification of (F_0, β) implies identification of each of the conditional choice probability functions

$$(7.1) \quad \phi_j(\xi) = \Pr(c = j \mid x = \xi), \quad j = 0, \dots, J.$$

relative to the collection of choice probability functions that satisfy all of the restrictions implicit in the random utility model.

The claimed identification follows from the following argument. One can derive several equivalent conditions that imply that the second alternative is chosen conditional on $x = \xi$.

$$(7.2) \quad \begin{aligned} c = 1 &\Leftrightarrow \bar{U}^{(j)} \leq \bar{U}^{(2)} \quad \text{for } j = 1, \dots, J+1 \\ &\Leftrightarrow \bar{\nu}^{(j)} - \bar{\nu}^{(2)} \leq \bar{V}^{(2)}(\xi, \beta) - \bar{V}^{(j)}(\xi, \beta) \quad \text{for } j = 1, \dots, J+1 \\ &\Leftrightarrow \nu^{(2)} \geq V^{(2)}(\xi, \beta) \quad \text{and} \quad \nu^{(j)} - \nu^{(2)} \leq V^{(j)}(\xi, \beta) - V^{(2)}(\xi, \beta) \\ &\quad \text{for } j = 1, 3, \dots, J+1. \end{aligned}$$

For each j ($1 \leq j \leq J$) form the matrix Γ_j by replacing each element of the j 'th column of a $J \times J$ identity matrix with -1 . (Let Γ_0 be the identity matrix itself.) Then the last condition in (7.2) is equivalent to $\Gamma_2 \nu \leq \Gamma_2 V(\xi, \beta)$.

Similarly $c = j \Leftrightarrow \Gamma_j \nu \leq \Gamma_j V(\xi, \beta)$. Hence

$$(7.3) \quad \phi_j(\xi) = \int \mathbf{1}(\Gamma_j \nu \leq \Gamma_j V(\xi, \beta)) \, dF_0(\nu).$$

So knowledge of F_0 and β is sufficient to derive all of the conditional choice probabilities.

The situation is illustrated in Figure 1 for the case $J = 2$. Given $x = \xi$, the point $V(\xi, \beta)$ partitions \mathbb{R}^2 into the three (almost) disjoint regions $\{ v: \Gamma_j v \leq \Gamma_j V(\xi, \beta) \}$, $j = 0, 1, 2$, which are labelled accordingly in the Figure. Alternative j is chosen if and only if the unobserved vector v takes a value in region j . Clearly the probabilities of these events can be determined from knowledge of β and F_0 .

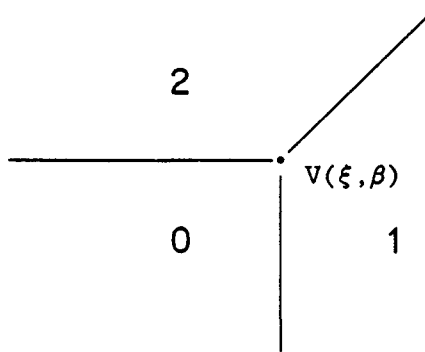


Figure 1.

The result holds despite the partial observability nature of the model. At first glance the result is surprising since intuition suggests that little can be learned about the preferences between alternatives 1 through J when one can only observe whether or not alternative 0 was chosen. Alternatively, the result can be interpreted as suggesting that unreasonably strong prior information is needed to completely identify (F_0, β) when actual choice behavior is only partially observed.

The result also implies that when one can observe some or all of the other y_j , $1 \leq j \leq J$, then the semiparametric random utility model is over-identified, given the sufficient conditions in Section 5. Alternatively, conditions that are not sufficient for identification given

partial observability may be sufficient when one can observe more about the actual choices. This last possibility is explored further in the next section.

8. IDENTIFICATION UNDER GENERAL OBSERVABILITY CONDITIONS

Here we extend the analysis of the previous sections by assuming that additional information is available on actual choices beyond that contained in y_0 . In particular, we assume that one can observe each of the random variables $x, y_0, \dots, y_M, 1 \leq M \leq J$. Hence we assume that the actual choice is observed if one of the first $M+1$ alternatives is chosen, but that no other information on the actual choice is observed. Full observability is the special case where $M = J$ or $M = J-1$. For $M \leq J-2$ this setting corresponds to data where two or more of the possible outcomes are aggregated into an "other" category by the data collection process.

The analysis proceeds by noting that each of the subproblems where only x and y_j is observed is equivalent to the problem studied in the previous sections. Hence the identification results that have already been developed are easily extended to cover the more general case.

Let the matrices $\Gamma_j, 0 \leq j \leq J$, be defined as before. For each j let F_j denote the distribution function of $\Gamma_j \nu$ under \mathbb{P} and let $V_j = \Gamma_j V$. Then for each alternative j we have

$$\begin{aligned}
 (8.1) \quad \phi_j(x) &= \Pr(\Gamma_j U \geq 0 \mid x) = \Pr(\Gamma_j \nu \leq V_j(x, \beta) \mid x) \\
 &= F_j(V_j(x, \beta)) \text{ (a.s.)}.
 \end{aligned}$$

This generalizes equation (2.3) to show that all of the conditional choice

probabilities have exactly the same structure provided that the appropriate transformation of U , V and ν is taken.

Suppose that it has been established that for each collection of distribution functions $\tilde{F}_j \in \mathcal{F}$ and each $b \in B$, for $0 \leq j \leq M$,

$$(8.2) \quad \tilde{F}_j(V_j(x,b)) = F_j(V_j(x,\beta)) \iff V_j(x,b) = H_j(V_j(x,\beta)) \text{ (a.s.)}$$

for some collection H_j of isotonic functions. That is, each of the random vectors $V_j(x,\beta)$ is individually identified up to an isotonic transformation. Then the constraints across the choice probability functions implied by the definitions of the V_j and F_j provide additional identifying power for β and F_0 .

For example, in the linear in parameters model we have

$V_j(x,b) = \Gamma_j b x = V(x,b_j)$ where $b_j = \Gamma_j b$. Let $\beta_j = \Gamma_j \beta$. Suppose that β_j is not identified relative to b_j only if $b_j = H_j \beta_j$ for some isotonic matrix H_j . The latter holds if and only if $b = \Gamma_j^{-1} H_j \Gamma_j \beta$. (In particular, for $j = 0$ we have $b = H_0 \beta$, since $\Gamma_0 = I$.) If this holds for $j \leq M$ and if β has full row rank then $\Gamma_j^{-1} H_j \Gamma_j = H_0$ for each $j \leq M$. This gives linear restrictions on the H_j that can be used to substantially reduce the set of isotonic matrices in which H_0 can lie.²⁴

Proposition 6 and its Corollary use this approach to generalize the identification results of Proposition 5.

Proposition 6: Suppose that Conditions (C1), (C4) and (C5) hold.

Suppose that for some collection of distribution functions $\tilde{F}_j \in \mathcal{F}$ and some $b \in B$ we have $\tilde{F}_j(V_j(x,b)) = F_j(V_j(x,\beta))$ (a.s.) for $0 \leq j \leq M$ ($M \geq 1$). Then there is an isotonic matrix H such that $b = H\beta$ and $\tilde{F}_j = F_j \circ H^{-1}$, $0 \leq j \leq M$. H has the form $H = \alpha P$, where $\alpha > 0$ is scalar and where P is a permutation

matrix for which the first M rows coincide with the identity.

Proposition 6, Corollary: Suppose that the conditions of Proposition 6 are met for $M = J-1$ or $M = J$. Then there is a scalar $\alpha > 0$ such that $b = \alpha\beta$ and $\tilde{F}_j = F_j \circ (\alpha I)^{-1}$, $j = 0, \dots, J$.

Proposition 6 establishes conditions under which β is identified up to a positive scalar transformation and a permutation of the rows corresponding to the unobserved y_j . As in the case where only y_0 is observed, one must normalize against permutations of the last $J-M$ rows of β . The need arises because for $j, k > M$, the events $\{c = j\}$ and $\{c = k\}$ are observationally equivalent.

Intuition suggests that there should be approximately a one-to-one correspondence between the number of y_j that are not observed and the number of scale normalizations needed on the model. The Proposition shows that this intuition is incorrect. Only a single scale normalization is needed to identify the scale of β when any two of the y_j are observed. So when $M = 0$ one needs J scale normalizations, but when $M \geq 1$ a single normalization suffices.

This apparent anomaly can be explained as follows. Consider the rescaling of utility differences DU , where $D = \text{diag}(\alpha_1, \dots, \alpha_j)$, where $\alpha_j > 0$, $j \leq J$. Since

$$(8.3) \quad \tilde{U}^{(1)} - \tilde{U}^{(j+1)} \geq 0 \iff \alpha_j \tilde{U}^{(1)} - \alpha_j \tilde{U}^{(j+1)} = \alpha_j U^{(j)} \geq 0$$

the probability that the first alternative is preferred to each of the others is not affected.

But consider any other pair of alternatives (i, j) . Given U we know

that alternative j is preferred to alternative i if and only if

$$(8.4) \quad (\Gamma_j U)^{(i)} = \tilde{U}^{(j+1)} - \tilde{U}^{(i+1)} \geq 0.$$

In order for the the transformation from U to DU to leave the implied ranking of these alternatives unchanged we must have $(\Gamma_j U)^{(i)} \geq 0 \Leftrightarrow (\Gamma_j DU)^{(i)} \geq 0$.

But the latter inequality can be written

$$(8.5) \quad (\Gamma_j DU)^{(i)} = (\alpha_i - \alpha_j) \tilde{U}^{(1)} + \alpha_j \tilde{U}^{(j+1)} - \alpha_i \tilde{U}^{(i+1)} \geq 0.$$

In order for (8.4) and (8.5) to be equivalent we must have $\alpha_i = \alpha_j$. Since $y_j = 1$ if and only if alternative j is preferred to all of the others, the conditional distribution of y_j is unaffected by the rescaling if and only if all of the α_j coincide.

In retrospect this fact is less surprising. A rescaling of the utility differences U is consistent with a rescaling of the actual utilities \tilde{U} if and only if each component of U is scaled by the same factor. This shows that the extreme case $M = 0$ of partial observability is the anomalous case. The relative scale of the rows of β usually will be identified.

Department of Economics
University of Minnesota
Minneapolis, MN 55455.

APPENDIX

This Appendix provides proofs for each of the propositions and corollaries in the text. Several of the proofs refer to the following lemmas.

Lemma A: Suppose that F is a proper distribution function on \mathbb{R}^J . Then $F(t+s) = F(t)$ for every $t \in \mathbb{R}^J$ only if $s = 0$.

Proof: Suppose that $s \neq 0$. If $F(t) = F(t+s)$ for all $t \in \mathbb{R}^J$, then it must also be true that $F(t) = F(t-s)$ for all $t \in \mathbb{R}^J$. (Set $t = \bar{t} - s$ for any arbitrary point $\bar{t} \in \mathbb{R}^J$.) So there is no loss of generality in assuming that $s^{(1)} < 0$. Choose some t_0 such that $F(t_0) > 0$, and construct a sequence t_i by iterating according to $t_{i+1} = t_i + s$. Then $F(t_i) = F(t_0)$ for all i , and $t_i^{(1)} \rightarrow -\infty$. Then

$$(A.1) \quad 0 < F(t_0) - F(t_i) \leq \int 1(v^{(1)} \leq t_i^{(1)}) dF(v)$$

for each i . But since F is a proper probability distribution the limit of the right hand side of (A.1) as $i \rightarrow \infty$ is zero. Since (A.1) implies a contradiction, the supposition that $s \neq 0$ must be false. ■

Lemma B: Fix $j \leq J+1$ and let Γ_j be the corresponding matrix defined in Section 7. Suppose that G and H are isotonic matrices for which $G = \Gamma_j^{-1} H \Gamma_j$. Then $H_{jj} > 0$ and there is a permutation matrix P with $P_{jj} = 1$ such that $G = H = H_{jj} P$.

Proof: By Proposition 3 there is exactly one strictly positive element in each row and in each column of G and H. For each i let h_i denote the strictly positive element in the i'th row of H. One can verify directly that $\Gamma_j^{-1} = \Gamma_j$, so $G = \Gamma_j H \Gamma_j$. Then direct matrix multiplication shows that for each i, k

$$(A.2) \quad G_{ik} = \left\{ \begin{array}{ll} -H_{jk} & \text{if } i = j \neq k \\ h_j & \text{if } i = j = k \\ h_j - h_i & \text{if } i \neq j = k \\ H_{ik} - H_{jk} & \text{if } i \neq j, k \neq j \end{array} \right\}.$$

Since H_{jk} and G_{jk} must both be non-negative we have $G_{jk} = H_{jk} = 0$ whenever $k \neq j$. Therefore every element in the j'th row of G and H except for G_{jj} and H_{jj} are zero. Since row j must have one positive element, $G_{jj} > 0$ and $H_{jj} > 0$. Then by definition of h_j we have $G_{jj} = h_j = H_{jj}$. Furthermore, since there can be at most one strictly positive element in each column of G and H we have $G_{ij} = H_{ij} = 0$ for $i \neq j$. Since $0 = G_{ij} = h_j - h_i$ for $i \neq j$, it follows that for each i, $h_i = h_j = H_{jj}$.

This establishes that all of the strictly positive elements of H take the same value H_{jj} . By Proposition 3, $H = PD$ for some permutation matrix P, where D is a diagonal matrix whose diagonal elements are given by the h_i . So $D = H_{jj} I$, and $H = PH_{jj} I = H_{jj} P$. Furthermore, since $H_{jj} = h_j > 0$ we must have $P_{jj} = 1$.

To complete the proof we need to establish that $G = H$. We have already shown that $G_{jk} = H_{jk} = 0$ whenever $k \neq j$, that $G_{ij} = H_{ij} = 0$ whenever $i \neq j$, and that $G_{jj} = H_{jj}$. So the only remaining detail is to show that $G_{ik} = H_{ik}$ whenever $i \neq j$ and $k \neq j$. But since $H_{jk} = 0$ whenever $k \neq j$, for these

elements we have $G_{ik} = H_{ik} - H_{jk} = H_{ik}$. ■

We now prove each of the propositions and corollaries in the text.

Proof of Proposition 1: Under the stated conditions

$$(A.3) \quad \begin{aligned} F(V(x,b)) &= \Pr(H(\nu) \leq H(V(x,\beta)) \mid x) \\ &= \Pr(\nu \leq V(x,\beta) \mid x) = F_0(V(x,\beta)), \end{aligned}$$

almost surely. ■

Proof of Corollary 1: Equation (3.1) holds by assumption. Apply the definition of an isotonic function to verify (3.2). ■

Proof of Corollary 2: Since the identity function is isotonic, apply Proposition 1 with $H = I$. ■

Proof of Corollary 3: By assumption we have

$V(x,b) = G(V(x,\beta)) + CV(x,\beta)$. Since the last term is almost surely zero, (3.1) and (3.2) hold for $H = G + \tilde{v}$. ■

Proof of Corollary 4: Conditions (3.5) - (3.7) imply that for each b in some open neighborhood of β (iii) continues to hold, and also

$$(A.4) \quad V^{(j)}(x_s, b) \leq V^{(j)}(x_t, b) \iff V^{(j)}(x_s, \beta) \leq V^{(j)}(x_t, \beta),$$

for each $j \leq J$, and each $s, t \leq n$. Fix any such $b \neq \beta$. Then for each $j \leq J$ there are infinitely many strictly increasing functions $h_j: \mathbb{R} \rightarrow \mathbb{R}$ such that

$$(A.5) \quad V^{(j)}(x_s, b) = h_j(V^{(j)}(x_s, \beta)), \quad s \leq n.$$

For each $j \leq J$ fix one of these functions and let $H: \mathbb{R}^J \rightarrow \mathbb{R}^J$ be defined by the identities $H^{(j)}(v) = h_j(v^{(j)})$, $j \leq J$. It is easy to verify that H is isotonic, and that (3.1) and (3.2) hold for this H . If (3.8) or (3.9) hold then it is possible to choose the h_j such that H does not coincide with the identity on the support of ν . ■

Proof of Corollary 5: Under the stated conditions $U \geq 0 \iff HU \geq 0$ (a.s.). Then (3.2) follows from the linearity of H and the definition of U . Equation (3.1) holds by assumption. F_0 is not identified relative to F , the distribution function for $H\nu$. But $\nu = H\nu$ if and only if $c'\nu = \nu$. ■

Proof of Proposition 2: We first prove that H is nonsingular. Let s be any vector such that $Hs = 0$. Then for each $t \in \mathbb{R}^J$ we have

$$(A.6) \quad F_0(t+s) = F(H(t+s)) = F(Ht) = F_0(t).$$

Then $s = 0$ by Lemma A, so H is nonsingular.

We next prove that $u \leq v \Rightarrow Hu \leq Hv$. If this is not true then there exist distinct points $u, v \in \mathbb{R}^J$ such that $u \leq v$ and $H^{(j)}u > H^{(j)}v$ for some j . Since H is nonsingular, without loss of generality assume that $u < v$. For $i = 1, 2, 3, \dots$ let $t_i = u + i(v-u)$. Clearly $t_i \rightarrow \infty$.²⁵ Furthermore, since $Ht_i = Hu + i(Hv - Hu)$, $H^{(j)}t_i \rightarrow -\infty$. So²⁶

$$(A.7) \quad F_0(t_i) = F(Ht_i) = \int_{[-\infty, Ht_i]} dF \\ \leq \int_{\mathbb{R}^J} 1\{v^{(j)} \leq H^{(j)}t_i\} dF(v) \rightarrow 0.$$

where the limit follows because the sequence of sets $\{v: v^{(j)} \leq H^{(j)}t_i\}$ converges from above to the empty set and because distribution functions are

continuous from above.²⁷

Since $t_i \rightarrow \infty$, it follows from the monotonicity of distribution functions that $F_0(t) = 0$ for all $t < \infty$. But then F_0 cannot be a proper distribution function, a contradiction. Hence $u \leq v \Rightarrow Hu \leq Hv$.

To prove the converse, note that the conditions of the Proposition are unchanged if one replaces H with H^{-1} and interchanges F with F_0 . Therefore the preceding argument also establishes that $u \leq v \Rightarrow H^{-1}u \leq H^{-1}v$. Replace u with Hu and v with Hv to establish the needed converse. ■

Proof of Proposition 3: Let \mathcal{H} be the set of matrices of the form given in the Proposition. Then $H \in \mathcal{H}$ if and only if its columns define a set of vectors that can be rescaled to coincide with the canonical set of basis vectors for \mathbb{R}^J . Hence $H \in \mathcal{H}$ if and only if the convex cone $\{Hv: z \geq 0\}$ generated by its columns coincides with the nonnegative orthant, that is if $z \geq 0 \Leftrightarrow Hz \geq 0$. But then $(\text{set } z = v - u) \mathcal{H}$ consists of those matrices for which $u \leq v \Leftrightarrow Hu \leq Hv$. ■

Proof of Proposition 4: Suppose $F(V(x,b)) = F_0(V(x,\beta))$ almost surely. Given identification of β we must have $V(x,b) = V(x,\beta)$, hence $F(V(x,b)) = F_0(V(x,\beta))$ almost surely. This implies that $F = F_0$ almost everywhere on V_0 . Since distribution functions are continuous from above, we must have $F = F_0$ on $\text{int}(V_0)$, hence on S_0 , hence everywhere. ■

Proof of Proposition 5: By hypothesis we have $F(bx) = F_0(\beta x)$ almost surely. Let $bx = b_1x_1 + b_2x_2$ be the decomposition of bx conformable with the decomposition in Condition (C5). Since β_1 is non-singular, there exists a

unique $J \times J$ matrix H such that $b_1 = H\beta_1$. Hence $bx = H\beta x + Cx_2$, where $C = b_2 - H\beta_2$. So

$$(A.8) \quad F(H\beta x + Cx_2) = F_0(\beta x) \text{ (a.s.)}.$$

Condition (C5) implies that the support of the distribution of βx conditional on x_2 covers \mathbb{R}^J almost surely. Then since distribution functions are continuous from above,

$$(A.9) \quad F(Ht + Cx_2) = F_0(t) \text{ for every } t \in \mathbb{R}^J \text{ (a.s.)}.$$

Let s be an arbitrary point in the support of Cx_2 . Then $F_0 = F_s \circ H$, where $F_s(t) = F(t+s)$ is a distribution function. An application of Proposition 2 proves that H is isotonic. Therefore H^{-1} exists, and

$$(A.10) \quad F(Ht) = F_0(t + Dx_2) \text{ for every } t \in \mathbb{R}^J \text{ (a.s.)},$$

where $D = -H^{-1}C$. (To see this, replace each t in (A.9) with $t - H^{-1}Cx_2$.)

This establishes that $F_0(t + Dx_2)$ is almost surely constant for every $t \in \mathbb{R}^J$. Let $M = L - J$ be the dimension of x_2 . Condition (C4) implies that the support of x_2 is not contained in any proper affine subspace of \mathbb{R}^M .

Therefore there exists a collection of $M+1$ affinely independent vectors c_i in the support of x_2 . For each of these vectors let $s_i = Dc_i$. Then for each $i, j \leq M+1$ and every $t \in \mathbb{R}^J$, we have $F_0(t + s_i) = F_0(t + s_j)$.

Since t is arbitrary we may replace it with $t - s_j$. Hence for each $i, j \leq M$ and every $t \in \mathbb{R}^J$, $F_0[t + (s_i - s_j)] = F_0(t)$. Lemma A establishes that each of the vectors $s_i - s_j = 0$. So $D(c_i - c_{M+1}) = 0$ for $i = 1, \dots, M$. Since the set of vectors $(c_i - c_{M+1})$ span \mathbb{R}^M , conclude that $D = 0$.

We have established that $D = H^{-1}C = \beta_2 - H^{-1}b_2 = 0$. Therefore $b_2 = H\beta_2$. Since also $b_1 = H\beta_1$, we have $b = H\beta$. That $F = F_0 \circ H^{-1}$ follows immediately from $b = H\beta$, given that the support of βx covers \mathbb{R}^J and that H is

nonsingular.

The Corollary follows immediately from Propositions 3 and 5. ■

Proof of Proposition 6: Each of the matrices Γ_j is nonsingular. Therefore the Conditions (C1), (C4) and (C5) continue to hold when V is replaced with V_j . So by Proposition 5, $\tilde{F}_j(V_j(x,b)) = F_j(V_j(x,\beta))$ implies that $\Gamma_j b = H_j \Gamma_j \beta$, or equivalently that $b = \Gamma_j^{-1} H_j \Gamma_j \beta$, for some isotonic matrix H_j . Since (C5) implies that β has full row rank, we must have $\Gamma_j^{-1} H_j \Gamma_j = H_0$ for each $j \leq M$. Lemma B establishes that $\Gamma_j^{-1} H_j \Gamma_j = H_0$ only if $H_0 = H_j = \alpha P_j$, where $P_{jj} = 1$ and α is the common j 'th diagonal element of H_0 and H_j . The conclusion follows immediately when these conditions hold simultaneously for each $j \leq M$.

The Corollary follows from the fact that the only $J \times J$ permutation matrix that coincides with the identity on $J-1$ of its columns is the identity itself. ■

FOOTNOTES

¹ This research was supported in part by a summer research fellowship from the Graduate School of the University of Minnesota. Some of the results are based upon material in the author's dissertation at the University of Wisconsin. The author wishes to thank Arthur Goldberger, Hidehiko Ichimura, Lung-Fei Lee, Charles Manski, James Powell, Christopher Sims, and Harald Uhlig for helpful comments. They are absolved from any responsibility for errors committed in this paper.

² If Z is any collection of vectors or vector-valued functions then $Z^{(j)}$ denotes the collection of j 'th coordinates of the elements of Z .

³ In general $1(\cdot)$ denotes the binary indicator function for the event in braces.

⁴ The canonical partial ordering on \mathbb{R}^J is employed throughout this paper. For vectors v and t in \mathbb{R}^J the inequalities $v \leq t$ and $v < t$ are equivalent to the corresponding coordinatewise inequalities. That is, $v \leq t$ is equivalent to $v^{(j)} \leq t^{(j)}$ for every j , and $v < t$ is equivalent to $v^{(j)} < t^{(j)}$ for every j . Similarly, $v \geq t$ (resp. $v > t$) if and only if $t \leq v$ (resp. $t < v$).

⁵ This extension of \mathcal{F} is composed of all distribution functions corresponding to measures μ on $\bar{\mathbb{R}}^J$ that satisfy $\mu(\bar{\mathbb{R}}^J) = 1$. Some of these correspond to improper probability measures on \mathbb{R}^J . By extending the domain of functions in \mathcal{F} to the compact set $\bar{\mathbb{R}}^J$, we ensure that the parameter space is compact when endowed with a topology corresponding to the weak topology on the underlying probability measures.

6 McFadden did not impose the constraint $\|\theta\| = 1$, but instead placed a scale normalization on the univariate marginal distributions of ν .

7 The propositions and corollaries are proved in the Appendix.

8 A vector-valued function H is strictly increasing in each argument if $u \leq v$ and $u \neq v$ implies $H(u) \leq H(v)$ and $H(u) \neq H(v)$. That is, if for some j $u^{(j)} < v^{(j)}$ then there is a k such that $H^{(k)}(u) < H^{(k)}(v)$.

9 Poirier imposed the scale normalizations by restricting the marginal distributions of F_0 . This is equivalent to our approach of leaving F_0 unrestricted and imposing the scale normalizations on V .

10 This fact justifies the use of the term "normalization" to discuss the necessary restrictions on V and B in the preceding discussion.

11 The Corollary is easily extended to cover cases where $C(V(x,\beta)) = \bar{v}$ (a.s.) for some nonlinear function C . Thus generally identification may fail if the support of $V(x,\beta)$ is contained in any nonlinear manifold of dimension less than J . The more general result is useful when the parameterizations of V and B admit nonlinear transformations of $V(x,\beta)$.

12 In fact, for the linear in parameters model Corollary 2 is obtained as the special case of Corollary 3 in which G is the identity and $\bar{v} = 0$.

13 The constant terms in b normalize the scale of $V(x,b)$, while the sign restrictions normalize against permutations.

14 These values are chosen for convenience. The example can be modified to accommodate any values of β_{12} and β_{21} for which $\text{rank}(\beta) = 2$ and $\beta \in B$.

15 Poirier attributes verification of this result to Takeshi Amemiya.

16 This analogy should not be taken too far, however. In the parametric model considered by Poirier there are no constraints on the parameter space involving more than one parameter at a time. In the semiparametric setting there are numerous inequality constraints linking the infinity of implicit scalar parameters. So it is not clear whether or not an infinite support for x is necessary in the semiparametric model.

17 The H defined in the Corollary cannot be isotonic since there are no strictly positive elements in $H^{(1)}$.

18 The proof of the Proposition requires that F_0 be a proper probability distribution on \mathbb{R}^J . This condition must be added to the Proposition when \mathcal{F} is extended to include improper probability distributions.

19 That is, there does not exist a vector $c \in \mathbb{R}^L$ and a scalar c_0 such that $\mathbb{P}\{c'x = c_0\} = 1$.

20 The linear independence restriction can be partially relaxed if there are known equality constraints on the columns of β .

21 For example, if x includes only personal characteristics of the decision maker then there may not be any justification for placing any constraints on β .

22 That is, one can always find an open neighborhood of β that does not include any permutation of the rows of β (except for β itself).

23 Note, however, that a restriction of the form $\beta_{ij} = 1$ implies a scale normalization and a sign restriction. It also normalizes against permutations that exchange row i with each row k for which it is known a priori that $\beta_{kj} \neq 1$. This implies that it is somewhat misleading to simply count constraint equations when interpreting these results.

24 Lemma B in the Appendix characterizes the nature of these restrictions.

25 That is, each element of t_i diverges to $+\infty$.

26 For $s, t \in \bar{\mathbb{R}}^J$, the notations $[s, t]$, (s, t) , etc. shall denote intervals of \mathbb{R}^J . That is $[s, t]$ is the Cartesian product across $j = 1, \dots, J$ of the intervals $[s^{(j)}, t^{(j)}]$, etc.

27 A function on $\bar{\mathbb{R}}^J$ is continuous from above if and only if it is right-continuous separately in each argument.

REFERENCES

- Cosslett, Stephen R. (1983): "Distribution-free maximum likelihood estimator of the binary choice model," *Econometrica* 51, 765-782.
- Daganzo, Carlos (1979): *Multinomial Probit: The Theory and Application to Demand Forecasting*. New York: Academic Press.
- Feinstein, Jonathan S. (1989): "The Safety Regulation of U.S. Nuclear Power Plants: Violations, Inspections, and Abnormal Occurrences," *Journal of Political Economy* 97, 115-154.
- Gallant, A. Ronald (1987): "Identification and consistency in semi-nonparametric regression," in Truman F. Bewley (ed.), *Advances in Econometrics: Fifth World Congress, Volume I*. New York: Cambridge University Press.
- Han, Aaron K. (1987): "Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator," *Journal of Econometrics* 35, 303-316.
- Ichimura, Hidehiko (1987): "Estimation of Single Index Models," Ph.D. dissertation, Department of Economics, Massachusetts Institute of Technology, Cambridge.
- Ichimura, Hidehiko and Lung-Fei Lee (1990): "Semiparametric Estimation of Multiple Index Models: Single Equation Estimation," in William A. Barnett, James Powell and George Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. New York: Cambridge University Press.

- Klein, Roger W. and Richard H. Spady (1987): "An Efficient Semiparametric Estimator for Discrete Choice Models," Economics Research Group, Bell Communications Research, Morristown, New Jersey.
- Maddala, G.S. (1983): *Limited-dependent and qualitative variables in econometrics*. New York: Cambridge University Press.
- Manski, Charles F. (1975): "Maximum score estimation of the stochastic utility model of choice," *Journal of Econometrics* 3, 205-228.
- (1985): "Semi-parametric analysis of discrete response: asymptotic properties of the maximum score estimator," *Journal of Econometrics* 27, 313-333.
- (1988): "Identification of Binary Response Models," *Journal of the American Statistical Association* 83, 729-738.
- Matzkin, Rosa L. (1988): "A nonparametric and distribution-free estimator for the binary choice and the threshold crossing models," Cowles Foundation Discussion Paper No. 889, Yale University, New Haven.
- (1990): "A nonparametric maximum rank correlation estimator," in William A. Barnett, James Powell and George Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, New York: Cambridge University Press.
- McFadden, Daniel (1974): "Conditional logit analysis of qualitative choice behavior," in Paul Zarembka (ed.), *Frontiers in Econometrics*, New York: Academic Press.
- (1976): "Quantal Choice Analysis: A Survey," *The Annals of Economic and Social Measurement* 5, 363-390.

- (1981): "Econometric Models of Probabilistic Choice," in Charles F. Manski and Daniel McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge: M.I.T. Press.
- (1982): "Qualitative Response Models," in Werner Hildenbrand (ed.), *Advances in Econometrics*, Cambridge: Cambridge University Press.
- (1984): "Econometric Analysis of Qualitative Response Models," in Zvi Griliches and Michael D. Intriligator (eds.), *Handbook of Econometrics*, Volume 2, Amsterdam: North-Holland.
- Poirier, Dale J. (1980): "Partial Observability in Bivariate Probit Models," *Journal of Econometrics* 12, 209-217.
- Powell, James L., James H. Stock and Thomas M. Stoker, "Semiparametric Estimation of Index Coefficients," Alfred P. Sloan School of Management Working Paper 1793-86, Massachusetts Institute of Technology.
- Robertson, Tim, F.T. Wright and R.L. Dykstra (1988): *Order Restricted Statistical Inference*, Chichester: John Wiley and Sons.
- Stoker, Thomas M. (1986): "Consistent Estimation of Scaled Coefficients," *Econometrica* 54, 1461-1481.
- (1990): "Equivalence of Direct, Indirect and Slope Estimators of Average Derivatives," in William A. Barnett, James Powell and George Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, New York: Cambridge University Press.