COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1175

ESTIMATION OF NONPARAMETRIC FUNCTIONS IN SIMULTANEOUS EQUATIONS MODELS, WITH AN APPLICATION TO CONSUMER DEMAND

Donald J. Brown and Rosa L. Matzkin

March 1998

Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand*

Donald J. Brown
Department of Economics, Yale University
and
Rosa L. Matzkin
Department of Economics, Northwestern University

March 1998

Abstract

We present a method for consistently estimating nonparametric functions and distributions in simultaneous equations models. This method is used to identify and estimate a random utility model of consumer demand. Our identification conditions for this particular model extend the results of Houthakker (1950), Uzawa (1971) and Mas-Colell (1977), where a deterministic utility function is uniquely recovered from its deterministic demand function.

1 Introduction

We present necessary and sufficient conditions for the identification of nonparametric primitive functions in simultaneous equations models. These conditions are necessary and sufficient for the identification of a random utility function from a distribution of consumer demand. Moreover, we propose a new method that can be used to estimate this random utility function.

The interrelationship among the variables in economic models is often described by a system of simultaneous equations. Given data on the observable endogenous and exogenous variables in the system, one is typically interested in estimating the functions and distributions describing the relationships among the observable and unobservable variables or in estimating the primitive functions of the system. For example, in a partial equilibrium model, where the observable variables are the equilibrium price and quantity, income of the consumers, and cost of inputs of the firms,

*We gratefully acknowledge support from the National Science Foundation and comments from J. Altonji, H. Bierens, R. Blundell, B.W. Brown, G. Chamberlain, L. Hansen, J. Heckman, J. Horowitz, H. Ichimura, A. Lewbel, D. McFadden, E. Maasoumi, C. Manski, B. Meyer, W. Newey, J. Park, J. Powell, J. Rust, R. Sickles, M. Stinchcombe, T. Stoker, F. Vella, Q. Vuong, and from the participants at the several seminars and conferences where we presented this work.

one may be interested in estimating the market demand and supply functions, or in estimating the utility functions of the consumers and the production functions of the firms.

Many results on simultaneous equations models have assumed that these functions are known up to a finite dimensional parameter. Extending ideas in Brown (1983), Roehrig (1988) analyzes the identification of simultaneous equations models that are not restricted to finite dimensional parameterizations. In the partial equilibrium example, Roehrig's conditions can be used to study the identification of the market demand and supply functions. We provide an extension of Roehrig's result that can be used to study the identification of the nonparametric primitive functions (e.g., production and utility functions) in simultaneous equations models.

The method that we propose to estimate nonparametric functions in simultaneous equation models is based on the Closest Empirical Distribution method. This method was introduced in Manski (1983) to develop estimators for finite dimensional parameters and nonparametric distributions of unobservable disturbances in nonlinear-in-parameters simultaneous equation models. In this method, an estimator for the parameters is obtained by minimizing the distance between the empirical and true distributions of the observable and unobservable variables. When, as we will require, the unobservable variables are assumed to be independent of the observable exogenous variables, an estimator can be obtained by minimizing the distance between the joint distribution of the exogenous variables and the product of their marginal distributions. Since in our model the parameters are functions, we minimize this distance over a set of nonparametric functions rather than over a finite dimensional set. To measure the distance between distributions, we propose using the bounded Lipschitz metric, which is equivalent to using the Prohorov metric, but easier to compute. We show that our estimator is strongly consistent. A different estimation method for nonparametric simultaneous equations models was proposed by Newey and Powell (1988) . Their nonparametric two-stage least-squares method is based on integral equations. It requires a conditional mean restriction on the disturbances, given instruments, and completeness of the conditional distribution of the dependent variable, given instruments.

As an application of our results, we consider a random utility model for demand analysis. Random utility models provide a theoretically consistent way of modeling randomness in consumer demand data. In these models, the randomness in the data is generated by heterogeneity of preferences, where the heterogeneity is due to unobservable characteristics. (See, for example, McElroy (1981).) Random utility models provide an alternative to commonly used methods where the randomness in the data is assumed to be generated by measurement errors. In these latter methods, the errors are added to the demand function generated by a nonrandom utility function. The nonrandom utility function may vary across the population but only as a function of observable characteristics. If the additive errors were part of a random utility specification, then in general they would not be homoskedastic, since they will be functionally dependent on income and prices (Brown and Walker (1989)). These random utility models can be seen as extensions of the Random Utility Model that is used to analyze choices over discrete sets of alternatives (McFadden (1974)).

Identifying and estimating the distribution of preferences over a population of consumers may be necessary, for example, to evaluate the distribution of demand that would be generated by a change in a tax structure. Or, the distribution of preferences over a population of consumers may be needed to evaluate the changes in the distribution of welfare that would be incurred by some policy change.

We show that from data on consumer demand one can estimate the nonparametric distribution of a random utility function. The location function of the distribution, denoted U^*, is assumed to be a concave and smooth function over the set of feasible commodity bundles, but it is otherwise unknown. The variation around the location function is generated by adding to $U^*(x)$ the term $\varepsilon' x$, where ε is unobservable. This specification implies that the distribution of the marginal utilities at any vector of commodities x and across a population of consumers has location $DU^*(x)$ and variation given by the distribution of ε, $DU^*(x)$ denotes the gradient of U at x. We do not assume that the function U^* or the distribution of ε possesses a parametric structure.

One might think of specifying a random coefficients model for the demand function, instead of for the utility function, and use demand data to estimate the nonparametric distribution of the coefficients (see, for e.g., the linear random coefficients structure used in Hahn (1994) to calculate consumer surplus). One of the problems associated with using this approach is that in random coefficient models for demand functions, it is not clear how to impose on the distribution of the coefficients the restrictions generated by utility maximization. It is also not clear in those models how one could recover the distribution of preferences from the distribution of demand. In contrast, our procedure provides both an estimator for a distribution of demand that satisfies the restrictions imposed by utility maximization and an estimator for the distribution of preferences.

We use our result for the identification of nonparametric functions in simultaneous equations models to derive necessary and sufficient conditions for the identification of U^* and the distribution of ε. Our assumptions are extremely weak. Assuming the last coordinate of ε equals one, and the value of U^* is known at one point, we show that a necessary and sufficient condition for the identification of both U^* and the distribution of ε is for any $U \neq U^*$ there exist x such that

$$\frac{U_{ij}(x)}{U_K(x)+1} \neq \frac{U^*_{ij}(x)}{U^*_K(x)+1}, \quad i,j = 1,2,...,K.$$

where $U_{ij}(x)$ and $U_K(x)$ denote the ij-th partial derivative and the derivative with respect to the K-th coordinate of x, respectively. We present a convenient way of normalizing the set of utility functions to guarantee that this identification condition is satisfied.

This identification result provides an extension of the results of Houthakker (1950), Uzawa (1972) and Mas-Colell (1977), where a deterministic utility function is uniquely recovered from a deterministic demand function.

Although we have not done it in this paper, it is easy to extend our results for the random utility model to the case where the function U^* depends on a vector of observable characteristics, and to allow the distribution of ε to depend on observable characteristics. This latter extension will require, however, that the observable characteristics be discrete random variables, corresponding to different consumer groups. The value of ε for each of the individuals in the same group is assumed to be drawn from a common distribution. If we specialize our results to the case where the function U^* is known up to the value of some finite dimensional parameter vector, then asymptotic properties of our estimator can be derived.

The outline of the paper is as follows. In the next section we present a simultaneous equations model and study its identification and estimation. In Section 3 we present a random utility model for consumer demand and apply to it the results presented in Section 2. All the proofs are contained in the Appendix.

2 The Model

We consider simultaneous equations model of the form

$$g(y, x, \varepsilon; U^*) = 0$$

where $y \in R^G$ denotes a vector of endogenous observable variables, $x \in R^K$ denotes a vector of exogenous observable variables, and $\varepsilon \in R^G$ denotes a vector of exogenous unobservable variables, which is distributed independently of x. The function $g : R^{G+K+G} \to R^G$ is assumed to be known up to the vector valued function $U^* : R^{G+K} \to R^S$. The distribution of ε is assumed to be unknown. We denote the cdf of ε by F_ε^* and the joint distribution of (x, ε) by $\Phi_{X,\varepsilon}^*$.

A simple example of a simultaneous equations model where g is known and U^* is unknown is a competitive equilibrium model with a representative firm, a representative consumer, and additive disturbances, where the unknown functions are the production function of the firm and the utility function of the consumer. Let $y = (p, q)$, $x = (w, I)$,and $\varepsilon = (\varepsilon_s, \varepsilon_d)$,where p and q denote, respectively, price and quantity, w and I denote, respectively, the price of the inputs paid by the firm and the income available to the consumer, and ε_s and ε_d denote unobservable disturbances in the price offered by the firm and the quantity demanded by the consumer. Denote the production function of the firm by T^* and the utility function of the consumer by V^*. Then,

$$g(y, x, \varepsilon; U^*) = \begin{pmatrix} g_1(y, x, \varepsilon; U^*) \\ g_2(y, x, \varepsilon; U^*) \end{pmatrix} = \begin{pmatrix} p - s(q, w; T^*) - \varepsilon_s \\ q - d(p, I; V^*) - \varepsilon_d \end{pmatrix}$$

where $s(q, w; T^*)$ denotes the profit maximizing price offered by a firm with production function T^* when it has to produce the quantity q and the price of the inputs is w, and $d(p, I; V^*)$ denotes the utility maximizing quantity demanded by a consumer with utility function V^* when the price of the commodity is p and the income of the consumer is I.

Let W denote the set to which U^* belongs, and let Γ denote the set to which the joint distribution $\Phi_{X,\varepsilon}^*$ belongs. Each $(U, \Phi_{X,\varepsilon}) \in (W \times \Gamma)$ generates through the system of structural equations $g(y,x,\varepsilon;U)=0$ a distribution, $\Psi_{Y,X}(\cdot;U,\Phi_{X,\varepsilon})$ of the observable variables.

As usual, we say that $(U^*,\Phi^*_{X,\varepsilon})$ is *identified* within the set $(W\times\Gamma)$ if

$$[(U,\Phi_{X,\varepsilon})\in(W\times\Gamma)\ \&\ (U,\Phi_{X,\varepsilon})\neq(U^*,\Phi^*_{X,\varepsilon})]\Rightarrow[\Psi_{Y,X}(\cdot;U,\Phi_{X,\varepsilon})\neq\Psi_{Y,X}(\cdot;U^*,\Phi^*_{X,\varepsilon})].$$

In words, $(U^*,\Phi^*_{X,\varepsilon})$ is identified within the set $(W\times\Gamma)$ if for any pair $(U,\Phi_{X,\varepsilon})$ that belongs to $(W\times\Gamma)$ and is different from $(U^*,\Phi^*_{X,\varepsilon})$, then the distribution of the observable variables (y,x) generated by $(U,\Phi_{X,\varepsilon})$ and the structural equations $g(y,x,\varepsilon;U)=0$ is different from that generated by $(U^*,\Phi^*_{X,\varepsilon})$ and the structural equations $g(y,x,\varepsilon;U^*)=0$. Equivalently, $(U^*,\Phi^*_{X,\varepsilon})$ is identified within the set $(W\times\Gamma)$ if from the distribution $\Psi_{Y,X}(\cdot;U^*,\Phi^*_{X,\varepsilon})]$ of observable variables, (y,x), one can uniquely recover the pair $(U^*,\Phi^*_{X,\varepsilon})$,within the set of functions and distributions $(W\times\Gamma)$ that satisfy $g(y,x,\varepsilon;U)=0$.

We next present a set of assumptions sufficient for establishing the identification of $(U^*,\Phi^*_{X,\varepsilon})$ within a set $(W\times\Gamma)$. For any $U\in W$, let $B_U=\{(y,x,\varepsilon)|g(y,x,\varepsilon;U)=0\}$.

ASSUMPTION I.0: $(U^*,\Phi^*_{X,\varepsilon})\in(W\times\Gamma)$.

ASSUMPTION I.1: $\forall U\in W$, *$g(\cdot;U)$ is C^1.*

ASSUMPTION I.2: $\forall U\in W$, *the matrices $\partial_y g(\cdot;U)$ and $\partial_\varepsilon g(\cdot;U)$ are of full rank.*

ASSUMPTION I.3: $\forall U\in W$, *there exist C^1 functions $r(y,x;U)$ and $h(x,\varepsilon;U)$ such that $\forall(y,x,\varepsilon)\in B_U$, $g(y,x,r(y,x;U);U)=0$ and $g(h(x,\varepsilon;U),x,\varepsilon;U)=0$.*

ASSUMPTION I.4: *The vectors of exogenous variables, x and ε, are stochastically independent.*

ASSUMPTION I.5: Γ *is the set of absolutely continuous distributions on R^{K+G} that possess everywhere positive Lebesgue densities.*

ASSUMPTION I.6: $\forall U,U'\in W$, $[(U\neq U^*)\iff(g(\cdot;U)\neq g(\cdot;U'))]$.

Theorem 1, which is a simple modification of Roehrig (1988), establishes necessary and sufficient conditions for the identification of $(U^*,\Phi^*_{X,\varepsilon})$.

THEOREM 1: *Suppose that Assumptions* I.0–I.6 *are satisfied. Then $(U^*,\Phi^*_{X,\varepsilon})$ is identified within $(W\times\Gamma)$ if and only if $\forall(U,\Phi_{X,\varepsilon})\in(W\times\Gamma)$ if $(U,\Phi_{X,\varepsilon})\neq(U^*,\Phi^*_{X,\varepsilon})$, then there exists $(y,x,\varepsilon,\varepsilon')$ such that $(x,\varepsilon)\in\mathrm{supp}(\Phi^*_{X,\varepsilon})$, $(x,\varepsilon)\in\mathrm{supp}(\Phi_{X,\varepsilon})$, $(y,x,\varepsilon)\in B_{U^*}$, $(y,x,\varepsilon')\in B_U$, and the rank of the matrix*

$$\begin{bmatrix}\partial_{y,x}g(y,x,\varepsilon;U) & \partial_{\varepsilon_i}g(y,x,\varepsilon';U)\\ \partial_{y,x}g(y,x,\varepsilon;U^*) & 0\end{bmatrix}\text{ is }2G$$

where ε_i denotes the vector ε with its i-th coordinate deleted.

As Roehrig (1988) notes, if $\forall i, j$ $(i \neq j)$ $\partial_{\varepsilon_j} g_i(\cdot; U) = 0$ and $\partial_{\varepsilon_i} g_i(\cdot; U) \neq 0$, the rank condition in Theorem 1 is equivalent to the condition that for some i the matrix

$$\begin{bmatrix} \partial_{y,x} r_i(y, x; U) \\ \partial_{y,x} r(y, x; U^*) \end{bmatrix} \text{ has rank } G + 1,$$

where $r_i(\cdot; U)$ denotes the i-th coordinate of the function $r(\cdot; U)$.

We provide in the following Theorem a condition equivalent to the rank condition of Theorem 1 that does not require calculating the rank of a matrix.

THEOREM 1′: *Suppose that Assumptions I.0–I.6 are satisfied. Then $(U^*, \Phi^*_{X,\varepsilon})$ is identified within $(W \times \Gamma)$ if and only if $\forall (U, \Phi_{X,\varepsilon}) \in (W \times \Gamma)$ if $(U, \Phi_{X,\varepsilon}) \neq (U^*, \Phi^*_{X,\varepsilon})$, then there exists $(y, x, \varepsilon, \varepsilon')$ such that $(x, \varepsilon) \in \text{supp}(\Phi^*_{X,\varepsilon})$, $(x, \varepsilon) \in \text{supp}(\Phi_{X,\varepsilon})$, $(y, x, \varepsilon) \in B_{U^*}$, $(y, x, \varepsilon') \in B_U$, and*

$$\partial_x h(x, \varepsilon; U^*) \neq \partial_x h(x, \varepsilon'; U).$$

To estimate nonparametric functions and distributions in simultaneous equations models satisfying the above assumptions, we propose using the Closest Empirical Distribution method. This method was introduced in Manski (1983) for parametric functions $r(\cdot; \theta)$, $\theta \in R^K$.

Given $U \in W$ and N independent observations $\{y^i, x^i\}_{i=1}^N$, let

$$\begin{aligned} F_{\varepsilon,X,N}(\eta, x; U) &= \frac{1}{N} \sum_{i=1}^{N} 1[\eta \geq r(y^i, x^i; U),\ x \geq x^i] \\ F_{X,N}(x) &= \frac{1}{N} \sum_{i=1}^{N} 1\left[x \geq x^i\right] \text{ and} \\ F_{\varepsilon,N}(\eta; U) &= \frac{1}{N} \sum_{i=1}^{N} 1[\eta \geq r(y^i, x^i; U)]. \end{aligned}$$

$F_{\varepsilon,X,N}(\eta, x; U)$ is the empirical joint distribution of the random vectors $r(Y, X; U)$ and X. $F_{X,N}(x)$ is the empirical distribution of the exogenous observable variables X, and $F_{\varepsilon,N}(\eta; U)$ is the empirical marginal distribution of the random vector $r(Y, X; U)$. Let $F_{\varepsilon,X}(\eta, x)$, $F_X(x)$ and $F_{\varepsilon,}(\eta; U)$ denote, respectively, the distributions of the random vectors $(r(\cdot; U), X)$, X, and $r(\cdot; U)$ induced by $\Psi_{Y,X}(\cdot; U^*, \Phi^*_{X,\varepsilon})$.

Let ρ denote the Prohorov metric on the space of distributions and let ρ_{BL} denote the bounded Lipschitz metric on the same space. The bounded Lipschitz distance between two distributions, F and G, is defined by

$$\rho_{BL}(F, G) = \sup_{\psi} \left| \int \psi(s) dF(s) - \int \psi(s) dG(s) \right|$$

where the supremum is taken over all functions ψ satisfying the Lipschitz condition:

$$|\psi(s) - \psi(s')| \leq m(s, s').$$

The metric $m(\cdot,\cdot)$ is any metric whose value is bounded by 1 on the joint domain of the distributions. For example, one can take m to be given by

$$m(s,s') = \frac{||s-s'||}{1+||s-s'||}.$$

The Prohorov metric and the bounded Lipschitz metrics are topologically equivalent (Corollary 4.3 in Huber (1981)). The bounded Lipschitz metric is, however, easier to calculate. The proof of Theorem 4.2 in Huber (1981) and the definitions of $F_{\varepsilon,X,N}(\cdot;U)$, $F_{\varepsilon,N}(;U)$, and $F_{X,N}$ imply that $\rho_{BL}(F_{\varepsilon,X,N}(\cdot;U), F_{\varepsilon,N}(;U)F_{X,N})$ is the optimal value of the objective function of the following linear programming problem:

$$\begin{aligned}
&\min_{\{u_{ijk}\}} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} m((\varepsilon_i,x_i),(\varepsilon_j,x_k)) \;\; \mu_{ijk} \\
\text{s.t.} \quad & \sum_{i=1}^{N} \mu_{ijk} = \frac{1}{N^2}, \quad j,k=1,...,N \\
& \sum_{j=1}^{N}\sum_{k=1}^{N} \mu_{ijk} = \frac{1}{N}, \quad i=1,...,N \\
& \mu_{ijk} \geq 0, \quad i,j,k=1,...,N.
\end{aligned}$$

We define our estimator, $\hat{U}_N$, for U^* to be any function that solves the following problem:

$$\min_{U\in W} \; \rho_{BL}(F_{\varepsilon,X,N}(\cdot;U), F_{\varepsilon,N}(;U)F_{X,N}).$$

Our estimator, $\hat{F}_N$ for F_ε^* is $F_{\varepsilon,N}(\varepsilon;\hat{U}_N)$.

Our consistency result for the above estimators will make use of the following assumptions:

ASSUMPTION C.1: $(U^*, \Phi^*_{X,\varepsilon})$ *is identified within a set* $(W \times \Gamma)$.

ASSUMPTION C.2: *The set* W *is a compact set with respect to a metric* $d: W\times W \to R_+$, e.g., C^2 uniform convergence on compacta.

ASSUMPTION C.3: *The metric* d *is such that convergence with respect to* d *of any sequence* $\{U_k\} \subset W$ *to a function* $U \in W$ *implies that* $r(\cdot;U_k)$ *converges to* $r(\cdot;U)$ *in the topology of* C^0 *uniform convergence on compacta.*

THEOREM 2: *Suppose that Assumptions* I.2, I.4, *and* C.1–C.3 *are satisfied. Then,* $\hat{U}_N$ *is a strongly consistent estimator for* U^* *with respect to the metric* d *and* $\hat{F}_N$ *is a strongly consistent estimator for* F_ε^* *with respect to the metric* ρ.

Note that if F_ε^* is assumed to possess absolutely continuous marginal distributions, then the convergence, in Theorem 2, of $\hat{F}_N$ to F_ε^* is in fact with respect to the supremum norm. This is so because weak convergence of cdf's is equivalent to convergence with respect to the supremum norm when the marginal cdf's of the limiting function are absolutely continuous (see Rao (1962)).

3 A Random Utility Model

As an application of our general results for simultaneous equations models, we consider a model of consumer demand where a typical consumer has a utility function $V : X \to R$ defined by

$$V(x) = U^*(x) + \varepsilon' x,$$

where $X \subset R^K_{++}$, $U^* : X \to R$ and $\varepsilon = (\varepsilon_1, ..., \varepsilon_K) \in R^K_+$ is such that $\varepsilon_K = 1$. The price of the commodities is given by a vector $p = (p_1, ..., p_K) \in R^K_{++}$, which we normalize by requiring that $p_K = 1$. The consumer, facing prices p and income $I \in R_{++}$, purchases a commodity bundle x^* that maximizes V^* over the set $\{x \in X | p'x \leq I\}$.

We assume the function U^* is a smooth utility function, in the sense of Debreu; i.e., we assume

(i) the closure of the indifference curves of U^* lie in R^K_{++},

(ii) $\forall x \in X \quad DU^*(x) >> 0$ and

(iii) $\forall x \in X \quad D^2U^*(x)$ is negative definite on the kernel of $DU^*(x)$

where $DU^*(x)$ and $D^2U^*(x)$ denote, respectively, the gradient and Hessian of U^* at x. We note that these conditions on U^* suffice to guarantee the existence of a C^1 demand function, for any given value of ε.

The monotonicity of U^* (condition (ii)) guarantees that the demand, x^*, of each consume satisfies the budget equality: $p'x^* = I$. Hence, given (p, I) and noting that $p_K = 1$, we can write $x^* = (x_1^*, ..., x_{K-1}^*, x_K^*) = (x_1^*, ..., x_{K-1}^*, I - \Sigma_{j=1}^{K-1} p_j x_j^*)$.

While the value of ε is assumed to be fixed and known for each consumer, different consumers have possibly different values of ε. The econometrician does not observe ε. The distribution of $\varepsilon = (\varepsilon_1, ..., \varepsilon_{K-1})$ over the population of consumers is assumed to be characterized by an absolutely continuous distribution function $F^*_\varepsilon : R^{K-1} \to R$. Since the demand function of each consumer depend on his/her value of ε, the distribution of ε over the population of consumers generates, for each vector of prices and income (p, I), a distribution of demand. The most the econometrician can observe is the joint distribution of purchased commodity bundles x^*, prices p, and incomes.

Applying our identification result in Theorem 1, we next show that both the smooth utility function U^* and the distribution F^*_ε of ε can be recovered from the joint pdf of (x^*, p, I). For simplicity, we denote $(x_1^*, ..., x_{K-1}^*)$ by x, $(p_1, ..., p_{K-1})$ by p, and $(\varepsilon_1, ..., \varepsilon_{K-1})$ by ε. Let $f_{x,p,I}(x, p, I)$ denote the joint pdf of (x, p, I), and let $F(x, p, I)$ denote its distribution. Assume that U^* belongs to a set W of smooth utility functions and that F^*_ε belongs to a set Γ of absolutely continuous distributions on R^{K-1}_+.

For any pair $(U,F)\in(W\times\Gamma)$, one can derive the pdf for (x,p,I), generated by (U,f). We denote this pdf by $f_{x,p,I}(x,p,I;U,f)$. Its conditional pdf's, given (p,I), can be calculated by letting x be, for any given (p,I), the vector that maximizes the function $U(x,I-p\cdot x)+\varepsilon' x+I-p\cdot x$ over the set $\{x\in R_+^{K-1}|I-p\cdot x\geq 0\}$.

THEOREM 3: *Assume that the vector (p,I) has a continuous Lebesgue density. Let $\bar{x}\in X$ and $\alpha\in R$ be given. Suppose that W is a set of smooth utility functions $U:X\to R$ such that $\forall U\in W$, $U(\bar{x})=\alpha$. Let $\varepsilon_K=1$, and denote by Γ the set of absolutely continuous distribution functions of vectors $(\varepsilon_1,...,\varepsilon_{K-1})$ that are distributed independently of (p,I). Then, (U^*,F_ε^*) is identified in $(W\times\Gamma)$ if and only if $\forall U\in W$ such that $U\neq U^*$, there exists $x\in X$ and $i,k\{1,...,K\}$ such that*

$$\frac{U_{ik}(x)}{U_K(x)+1}\neq\frac{U^*_{ik}(x)}{U^*_K(x)+1}.$$

Note that the result of Theorem 3 can be used to determine not only when a nonparametric U is identified, but also when a parametric U is identified.

The following corollary present a set of conditions that are easily imposed and guarantee that the above identification conditions are satisfied.

COROLLARY 1: *Let $\bar{x}\in X$, $\alpha\in R, T\in R^K_{++}$, and $g:X\to R$ be known. Let W be a set of smooth utility functions $U:X\to R$ such that $\forall U\in W$,*

(i) $U(\bar{x})=\alpha$,

(ii) $DU(\bar{x})=T$, *and*

(iii) $\forall x\in X$, $U_K(x)=g(x)$.

Let $\varepsilon_K=1$, and denote by Γ the set of continuous density functions of vectors $(\varepsilon_1,...,\varepsilon_{K-1})$ that are distributed independently of (p,I). Then,

$$(U^*,F_\varepsilon^*)\text{ is identified in }(W\times\Gamma).$$

Estimators for U^* and F_ε^* can be obtained using the result in the previous section. Here, x, the commodity bundle, is the vector of endogenous variables, (p,I), the vector of prices and income is the vector of exogenous variables, and ε, the heterogeneity term, is the unobservable exogenous variables. For any function U in a set W to which U^* belong, we have that the first order conditions of the maximization over x of $U(x,I-p\cdot x)+\varepsilon\cdot x+I-p\cdot x$ subject to the constraints that $x\geq 0$, $I-p\cdot x\geq 0$ are

$$\varepsilon_k=r_k(x,p,I;U)=(U_K(x,I-p'x)+1)p_k-U_k(x,I-p'x),\quad k=1,...,K-1$$

Let W denote a set of smooth function that satisfy (i)–(iii) in Corollary 1 and is compact with respect to the metric d defined by

$$\begin{aligned}d(U,U') &= \sup_{x\in X}|U(x)-U'(x)|+\sup_{x\in X}\max_{1\leq k\leq K}|U_k(x)-U'_k(x)|\\ &\quad+\sup_{x\in X}\max_{1\leq j,k\leq K}|U_{kj}(x)-U'_{kj}(x)|\end{aligned}$$

where $U_k(x) = \partial U(x)/\partial x_k$ and $U_{kj}(x) = \partial^2 U(x)/\partial x_k \partial x_j$ $(k, j = 1, ..., K;\ U \in W;\ x \in X)$. W will be compact with respect to this metric if all functions in W and their derivatives up to the second order are equicontinuous and uniformly bounded. Then, in addition to Assumptions I.1–I.6 being satisfied we have that Assumptions C.1–C.3 are also satisfied. Hence, we can apply the consistency result in the previous section.

To compute the estimators, we note that the value of the function

$$Q_N(U) = \rho_{BL}(F_{\varepsilon,p,I,N}(\cdot;U)\ , F_{\varepsilon,N}(\cdot;U)\ F_{p,I,N}(\cdot))$$

at any function $U \in W$, depends on U only through the values that U and its gradients attain at the finite number of points $(x^1, I^1 - p^1 \cdot x^1), ..., (x^N, I^N - p^N \cdot x^N)$, we can transform the optimization problem (1) into a two step procedure. First, find the values and gradients of $\hat{U}_N$ at $(x^1, I^1 - p^1 \cdot x^1), ..., (x^N, I^N - p^N \cdot x^N)$, and then we obtain the function $\hat{U}_N$ by interpolating between the values. This follows the approach taken in Matzkin (1991, 1992).

For each vector $(u^1, ..., u^N, \partial u^1, ..., \partial u^N)$ of values and gradients of a function in W, let

$$\begin{aligned}\tilde{Q}_N(u^1, ..., u^N, \partial u^1, ..., \partial u^N) \ = \ & \rho_{BL}\left(\left(\tfrac{1}{N}\sum_{i=1}^{N} 1[(\varepsilon, p, I) \geq (\varepsilon^i, p^i, I^i)]\right),\right. \\ & \left.\times\left(\tfrac{1}{N}\sum_{i=1}^{N} 1[\varepsilon \geq \varepsilon^i]\right)\left(\tfrac{1}{N}\sum_{i=1}^{N} 1[(p, I) \geq (p^i, I^i)]\right)\right)\end{aligned}$$

where $\varepsilon^i_k = (\partial u^i_K + 1)p^i_k - \partial u^i_k$, $k = 1, ..., K$; $i = 1, ..., N$.

Then, the first step consists of finding a solution to the problem:

$$\min_{\{u^i\},\{\partial u^i\}} \tilde{Q}_N(u^1, ..., u^N, \partial u^1, ..., \partial u^N)$$

subject to

$$\begin{aligned} & u^i < u^j + \partial u^j \cdot ((x^i, I^i - p^i \cdot x^i) - (x^j, I^j - p^j \cdot x^j)), \quad i = 1, ..., N+1 \\ & u^{N+1} = \alpha, \quad \partial u^{N+1} = T \\ & \partial u^i_K = g(x^i, I^i - p^i \cdot x^i), \quad i = 1, ..., N+1 \end{aligned}$$

The constraints guarantees that the vectors $(u^1, ..., u^N, \partial u^1, ..., \partial u^N)$ over which the minimization is performed correspond to the values and gradients of a function in W.

The second step proceeds to interpolate the estimated values and gradients, $(\hat{u}^1, ..., \hat{u}^{N+1})$, and $\partial\hat{u}^1, ..., \partial\hat{u}^{N+1}$, of U^*.

The estimator $\hat{F}_\varepsilon$ for the cdf of ε is calculated by

$$\hat{F}_\varepsilon(\varepsilon) = \tfrac{1}{N}\sum_{i=1}^{N} 1[\varepsilon_k \geq ((\partial\hat{u}^i_K + 1)p^i_k - \partial\hat{u}^i_k);\ k = 1, ..., K-1].$$

4 Appendix

PROOF OF THEOREM 1: By Roehrig (1988), $r(\cdot;U^*)$ and $\Phi^*_{X,\varepsilon}$ are identified if and only if the rank condition is satisfied. Hence, by Assumption I.6, U^* and $\Phi^*_{X,\varepsilon}$ are identified if and only if the same rank condition is satisfied.

PROOF OF THEOREM 1$'$: By Roehrig (1988), the rank condition in the statement of Theorem 1 is equivalent to the condition that

$$\partial_x p(x,\varepsilon) \neq 0 \tag{1}$$

where the function p is defined by

$$p(x,\varepsilon) = r(h(x,\varepsilon;U^*),x;U). \tag{2}$$

Using (2), (1) becomes

$$\partial_y r(y,x;U)\partial_x h(x,\varepsilon;U^*) + \partial_x r(y,x;U) \neq 0. \tag{3}$$

The equations $g(y,x,\varepsilon;U^*) = 0$ and $g(y,x,\varepsilon';U) = 0$ imply, by Assumption I.2 and the Implicit Function Theorem that

$$\partial_y r(y,x;U) \; = \; -(\partial_\epsilon g(y,x;U))^{-1}\partial_y g(y,x;U) \text{ and} \tag{4}$$
$$\partial_x r(y,x;U) \; = \; -(\partial_\varepsilon g(y,x;U))^{-1}\partial_x g(y,x;U)\,. \tag{5}$$

Hence, (3) becomes

$$(\partial_\varepsilon g(y,x;U))^{-1}(\partial_y g(y,x;U)\partial_x h(x,\varepsilon;U^*) + \partial_x g(y,x;U)) \neq 0. \tag{6}$$

By Assumption I.2, (6) is equivalent to

$$\partial_y g(y,x;U)\partial_x h(x,\varepsilon;U^*) + \partial_x g(y,x;U) \neq 0, \tag{7}$$

which, by Assumption I.2 and the Implicit Function Theorem is equivalent to

$$\partial_x h(x,\varepsilon;U^*) \neq -(\partial_y g(y,x;U))^{-1}\partial_x g(y,x;U) = \partial_x h(x,\varepsilon;U). \tag{8}$$

Hence, identification holds if and only if

$$\partial_x h(x,\varepsilon;U^*) \neq \partial_x h(x,\varepsilon;U).$$

PROOF OF THEOREM 2: We will show that the assumptions in Theorem 1 in Manski (1983) are satisfied when the parameter space is (W,d) instead of $(\Theta,||\cdot||)$, and when the distance over the space of probability distributions is ρ instead of the supremum distance used in Manski. The strong consistency of $\hat{U}_N$ will then follow by the same arguments used in the proof of Theorem 1 in Manski (1983).

The compactness of the parameter space follows by Assumption C.1. Identification follows by noting that Assumption I.2 guarantees that the conditional distributions of $r(Y,X;U)$ given X are identical to the conditional distributions of $r(Y,X;U^*)$

given X if and only if the joint pdf of (Y,X) generated by U and the joint distribution of $(r(Y,X;U),X)$ equals the pdf of (Y,X) generated by U^* and the joint distribution of (ε,X). Since our Assumption C.1 rules out this latter equality, the joint distribution of $(r(Y,X;U),X)$ must be different from the joint distribution of (ε,X) if and only if $U \neq U^*$. This is the argument of Lemma 3.2 in Roehrig (1988). See also Brown (1983, Lemma 1).

We next show that

(I) If $\{U_k\} \subset W, U \in W$ are such that $d(U_k,U) \to 0$ and if $F'_{Y,X}$ is any distribution of (Y,X) then

$$\rho(F'_{\varepsilon,X}(\cdot;U_k), F'_{\varepsilon}(\cdot;U_k)\ F'_X) \to \rho(F'_{\varepsilon,X}(\cdot;U), F'_{\varepsilon}(\cdot;U)F'_X)$$

where for any $U \in W$, $F'_{\varepsilon,X}(\cdot;U'), F'_{\varepsilon}(\cdot;U')$, and F'_X denote, respectively, the distributions of $(r(\cdot;U'),X)$, $r(\cdot;U')$, and X generated by $F'_{Y,X}$.

Let $\{U_k\} \subset W, U \in W$ be such that $d(U_k,U) \to 0$. Let μ_{ε_k} and μ_{ε} denote, respectively, the probability measures of $F'_{\varepsilon}(\cdot;U_k)$ and $F'_{\varepsilon}(\cdot;U)$. Let $\mu'_{Y,X}$ denote the probability measure of $F'_{Y,X}$. By assumption, $d(U_k,U) \to 0$ implies that $r(\cdot;U_k)$ converges to $r(\cdot;U)$ in the topology of C^0 uniform convergence on compacta. It then follows that $\mu'_{\varepsilon_k} = \mu'_{Y,X} \circ r^{-1}(\cdot;U_k)$ converges weakly to $\mu'_{\varepsilon} = \mu'_{Y,X} \circ r^{-1}(\cdot;U)$. (See Theorem 3.3 in Rao (1962).) Similarly, $\mu'_{\varepsilon_k,X}$ converges weakly to $\mu'_{\varepsilon,X}$, where $\mu'_{\varepsilon_k,X}$ and $\mu_{\varepsilon,X}$, are, respectively, the probability measures of $F'_{\varepsilon,X}(\cdot;U_k)$ and $F'_{\varepsilon,X}(\cdot;U)),X)$. Hence, since weak convergence of probability measures implies weak convergence of the distribution function, which implies convergence with respect to the metric ρ, we have that

$$\begin{aligned}
& |\rho(F_{\varepsilon,X}(\cdot;U_k), F_{\epsilon}(\cdot;U_k)F_X) - \rho(F_{\varepsilon,X}(\cdot;U), F_{\varepsilon}(\cdot;U)F_X)| \\
\leq\ & |\rho(F_{\varepsilon,X}(\cdot;U_k), F_{\varepsilon}(\cdot;U_k)F_X) - \rho(F_{\varepsilon,X}(\cdot;U), F_{\varepsilon}(\cdot;U_k)F_X)| \\
& + |\rho(F_{\varepsilon,X}(\cdot;U), F_{\varepsilon}(\cdot;U_k)F_X) - \rho(F_{\varepsilon,X}(\cdot;U), F_{\varepsilon}(\cdot;U)F_X)| \\
\leq\ & |\rho(F_{\varepsilon,X}(\cdot;U_k), F_{\epsilon,X}(\cdot;U))| + |\rho(F_{\varepsilon}(\cdot;U_k), F_{\varepsilon}(\cdot;U))| \\
\to\ & 0
\end{aligned}$$

where the last inequality follows from the triangle inequality by noting that for any F, G, D, $\rho(F,G) \leq \rho(F,D) + \rho(D,G)$ and $\rho(F,D) \leq \rho(F,G) + \rho(G,D)$ imply that $-\rho(G,D) \leq \rho(F,G) - \rho(F,D) \leq \rho(G,D)$.

This shows (I).

Next, we show that

(II) If $\{U_k\}_{k=1}^{\infty} \subset W, U \in W$ are such that $d(U_k,U) \to 0$ and if $\{F_{Y,X,k}\}_{k=1}^{\infty}$ is a sequence of distributions such that $F_{Y,X,k}$ converges weakly to $F_{Y,X}$, then

$$\rho(F_{\varepsilon,X,k}(\cdot;U_k), F_{\varepsilon,k}(\cdot;U_k)F_{X,k}) \to \rho(F_{\varepsilon,X}(\cdot;U), F_{\varepsilon}(\cdot;U)F_X$$

where for any $U' \in W$, $F_{\varepsilon,X,k}(\cdot;U')$, $F_{\varepsilon,k}(\cdot;U')$, and $F_{X,k}$ denote, respectively, the distributions of $(r(\cdot;U'),X)$, $r(\cdot;U')$, and X generated by $F_{Y,X,k}$, and

$F_{\varepsilon,X}(\cdot;U)$, $F_\varepsilon(\cdot;U)$ and F_X denote, respectively, the distributions of $(r(\cdot;U),X)$, $r(\cdot;U)$, and X generated by $F_{Y,X}$ ($F_{Y,X}$ is the true distribution of (Y,X).)

By assumption, $d(U_k,U)\to 0$ implies that $r(\cdot;U_k)$ converges to $r(\cdot;U)$ in the topology of C^0 uniform convergence on compacta. Let $\mu_{Y,X,k}$ denote the probability measure of $F_{Y,X,k}$, $\mu_{\varepsilon_k,}$ denote the probability measure of $F_{\varepsilon,k}(\cdot;U_k)$, and μ_ε denote the probability measure of $F_\varepsilon(\cdot;U)$. Then, since $\mu_{Y,X,k}$ converges weakly to $\mu_{Y,X}$, it follows that $\mu_{\varepsilon_k,}=\mu_{Y,X,k}\circ r^{-1}(\cdot;U_k)$ converges weakly to $\mu_\varepsilon=\mu_{Y,X}\circ r^{-1}(\cdot;U)$. (See Theorem 3.3 in Rao (1962).) Similarly, $\mu_{\varepsilon_k,X}$ converges weakly to $\mu_{\varepsilon,X}$, where $\mu_{\varepsilon_k,X}$ and $\mu_{\varepsilon,X}$, are respectively, the probability measures of $F_{\varepsilon,X,k}(\cdot;U_k)$ and $F_{\varepsilon,X}(\cdot;U)$. Since weak convergence of probability measures implies weak convergence of the distribution functions, which implies convergence with respect to the metric ρ, we see that

$$\begin{aligned}
&|\rho(F_{\varepsilon,X,k}(\cdot;U_k),F_{\varepsilon,k}(\cdot;U_k)F_X)-\rho(F_{\varepsilon,X}(\cdot;U),F_\varepsilon(\cdot;U)F_X)|\\
\leq\ &|\rho(F_{\varepsilon,X,k}(\cdot;U_k),F_{\varepsilon,k}(\cdot;U_k)F_X)-\rho(F_{\varepsilon,X}(\cdot;U),F_{\varepsilon,k}(\cdot;U_k)F_X)|\\
&+|\rho(F_{\varepsilon,X}(\cdot;U),F_{\varepsilon,k}(\cdot;U_k)F_X)-\rho(F_{\varepsilon,X}(\cdot;U),F_\varepsilon(\cdot;U)F_X)|\\
\leq\ &\rho(F_{\varepsilon,X,k}(\cdot;U_k),F_{\varepsilon,X}(\cdot;U))+\rho(F_{\varepsilon,k}(\cdot;U_k)F_X,F_\varepsilon(\cdot;U)F_X)\\
\to\ &0.
\end{aligned}$$

This shows (II).

(I) and (II) imply that the continuity assumption in Theorem 1 in Manski (1983) is satisfied. Hence, using the same arguments as in the proof of that theorem, we get that

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

Now to show that $\hat{F}_N$ converges a.s. to F_ε^*, we note that

$$\mu_{Y,X,N}\circ r^{-1}(\cdot;\hat{U}_N)\to\mu_{Y,X}\circ r^{-1}(\cdot;U^*) \text{ weakly,}$$

since $\mu_{Y,X,N}\to\mu_{Y,X}$ weakly, and by Assumption C.3, $d(\hat{U}_N,U^*)\to 0$ implies that $r(\cdot;\hat{U}_N)\to r(\cdot;U^*)$ in the topology of C^0 uniform convergence in compacta. Hence, $\hat{F}_N$, which is the distribution function of $\mu_{Y,X,N}\circ r^{-1}(\cdot;\hat{U}_N)$, converges weakly to F_ε^*, which is the distribution function of $\mu_{Y,X}\circ r^{-1}(\cdot;U^*)$.

This completes the proof of the theorem.

PROOF OF THEOREM 3: Choose the K-th good as numeraire and express the first order condition for utility maximization subject to a budget constraint in terms of the inverse demand function. Letting $V^*(x,\varepsilon)$ denote the smooth utility function, $U^*(x)+\varepsilon' x$, and denoting the associated smooth random inverse demand function as $g(x,\varepsilon)$, it follows that

$$g_i(x,\varepsilon)=\frac{\dfrac{\partial V^*(x,\varepsilon)}{\partial x_i}}{\dfrac{\partial V^*(x,\varepsilon)}{\partial x_K}},\quad i=1,...,K-1.$$

Then, the first order condition for utility maximization are given by

$$g(x,\varepsilon) = p$$
$$p'x = I$$

where $p' = (p_1, ..., p_{K-1}, 1)$. Using the last equation to calculate x_K, we can express the above system as:

$$\frac{U_i^*\left(x_1, ..., x_{K-1}, I - \sum_{j=1}^{K-1} p_j x_j\right) + \varepsilon_i}{U_K^*\left(x_1, ..., x_{K-1}, I - \sum_{j=1}^{K-1} p_j x_j\right) + 1} = p_i, \quad i = 1, ..., K-1.$$

Rearranging, we get, for $i = 1, ..., K-1$:

$$\varepsilon_i = \left[U_K^*(x_1, ..., x_{K-1}, I - \sum_{j=1}^{K-1} p_j x_j) + 1\right] p_i - U_i^*\left(x_1, ..., x_{K-1}, I - \sum_{j=1}^{K-1} p_j x_j\right).$$

This is a system of $K-1$ simultaneous equations, with $x_1, ..., x_{K-1}$ being the $K-1$ observable endogenous variables, $\varepsilon_1, ..., \varepsilon_{K-1}$ being the $K-1$ unobservable exogenous variables, and $p_1, ..., p_{K-1}, I$ denoting the K observable exogenous variables in the system. Denote the vectors $(x_1, ..., x_{K-1})$, $(\varepsilon_1, ..., \varepsilon_{K-1})$, and $(p_1, ..., p_{K-1})$ by x, ε, and p, respectively.

For each $U \in W$ define the function $r : X \times P \times R_{++} \to R^{K-1}$ by

$$r_i(x, p, I; U) = \left[U_K\left(x_1, ..., x_{K-1}, I - \sum_{j=1}^{K-1} p_j x_j\right) + 1\right] p_i - U_i\left(x_1, ..., x_{K-1}, I - \sum_{j=1}^{K-1} p_j x_j\right)$$

$i = 1, ..., K-1$. Our smoothness assumptions on the functions in W imply that:

(i) $\forall U \in W$, $r(\cdot; U)$ is a C^1 function, and

(ii) $\forall U \in W$, there exists a function $\chi : P \times R_{++} \times R_+^{K-1} \to R^{K-1}$ such that $\forall \varepsilon$ $\varepsilon = r(\chi(p, I, \varepsilon; U), p, I; U)$.

Moreover, one also has that

(iii) $\forall U, U' \in W \quad U \neq U' \Longleftrightarrow r(\cdot; U) \neq r(\cdot; U')$.

To show this last assertion, we first note that $U = U' \Rightarrow r(\cdot; U) = r(\cdot; U')$. Next, suppose that $U \neq U'$, then, since $U(\bar{x}) = U'(\bar{x})$, there must exist x and $i \in \{1, ..., K\}$ such that $U_i(x) \neq U_i'(x)$. If $i = K$ and for some $k \neq K$, $U_k(x) = U_k'(x)$, then letting p and I be such that $x_K = I - \sum_{j=1}^{K-1} p_j x_j$, we get that $r_k(x, p, I; U) \neq r_k(x, p, I; U')$. If $i = K$ and for some $k \neq K$, $U_k(x) \neq U_k'(x)$, let $p_k \neq (U_k(x) - U_k'(x))/(U_K(x) - U_K'(x))$, and choose p_j $(j \neq k)$ and I such that $x_K = I - \sum_{j=1}^{K-1} p_j x_j$. Then, $r_k(x, p, I; U) \neq$

$r_k(x,p,I;U')$. Suppose, on the other hand, that for all x, $U_K(x) = U'_K(x)$, then letting p and I be such that $x_K = I - \sum_{j=1}^{K-1} p_j x_j$, we get that $r_i(x,p,I;U) \neq r_i(x,p,I;U')$.

Properties (i)–(iii) allow us to use Theorem 1 to determine the necessary and sufficient conditions for the identification of U^* and the distribution F^*_ε of ε. Hence, to prove the theorem it only remains to show that for all $U \in W$ such that $U \neq U^*$, there exist vectors x, p, I and $i \in \{1,...,K-1\}$ such that the matrix

$$A_i = \begin{bmatrix} \dfrac{\partial r_i(x,p,I;U)}{\partial(x,p,I)} \\ \dfrac{\partial r_i(x,p,I;U^*)}{\partial(x,p,I)} \end{bmatrix}$$

has rank K.

The typical elements of the above matrix are as follows:

$$\begin{aligned} \frac{\partial r_i}{\partial x_j} &= U_{Kj}p_i - U_{KK}p_jp_i + U_{iK}p_j - U_{ij} \\ \frac{\partial r_i}{\partial p_j} &= U_{iK}x_j - U_{KK}x_jp_i \quad \text{if } i \neq j \\ \frac{\partial r_i}{\partial p_i} &= U_{iK}x_i - U_{KK}x_ip_i + U_K + 1 \\ \frac{\partial r_i}{\partial I} &= U_{KK}p_i - U_{iK} \end{aligned}$$

with U being replaced by U^* when the derivatives are taken over $r(\cdot;U^*)$.

Write the matrix of partial derivatives of $r(\cdot;U)$ and $r(\cdot;U^*)$ as a vector of columns

$$A = \begin{bmatrix} \dfrac{\partial r(x,p,I;U)}{\partial(x,p,I)} \\ \dfrac{\partial r_i(x,p,I;U^*)}{\partial(x,p,I)} \end{bmatrix} = [\partial x_1 \ \cdots \ \partial x_{K-1} \ \ \partial p_1 \ \cdots \ \partial p_{K-1} \ \ 1].$$

The rank of this matrix is the same as the rank of the matrix obtained by substituting each column ∂x_j with the column $(\partial x_j)^* = \partial x_j + p_j \ \ \partial I$ and substituting each column ∂p_j with $(\partial p_j)^* = \partial p_j + x_j \ \ \partial I$. The i-th and the$(K-1)+i$ elements of the vector $(\partial x_j)^*$ are then, respectively, $U_{Kj}p_j - U_{ij}$ and $U^*_{Kj}p_j - U^*_{ij}$. The i-th and the $(K-1)+i$ elements of the vector $(\partial p_j)^*$ are $U_K + 1$ and $U^*_K + 1$ if $i = j$ and they are 0 otherwise.

Substitute the $(\partial p_j)^*$ columns with the columns $(\partial p_j)^{**}$ obtained by dividing the $(\partial p_j)^*$ columns by $(U_K + 1)$. Subtract from each $(\partial x_j)^*$ column the vector $\sum_{i=1}^{K-1} (\partial x_j)^*_{K-1+i} \ \ (\partial p_i)^{**}$, where $(\partial x_j)^*_{K-1+i}$ is the $(K-1+i)$-th element of the column vector $(\partial x_j)^*$. Finally, subtract from the ∂I vector the vector $\sum_{i=1}^{K-1} (\partial I)_{K-1+i}$ $(\partial p_i)^{**}$. Then, the resulting matrix is

$$\begin{bmatrix} \hat{X} & \hat{P} & \hat{I} \\ \hat{X}^* & \hat{P}^* & \hat{I}^* \end{bmatrix}$$

where the submatrices $\hat{X}^*$ and $\hat{I}^*$ are zero matrices, and the submatrix $\hat{P}^*$ is the $(K-1)\times(K-1)$ identity matrix; i.e.,

$$\begin{bmatrix} \hat{X} & \hat{P} & \hat{I} \\ \hat{X}^* & \hat{P}^* & \hat{I}^* \end{bmatrix} = \begin{bmatrix} \hat{X} & \hat{P} & \hat{I} \\ 0 & I_{K-1} & 0 \end{bmatrix}$$

The ij-th element of $\hat{X}$ is $U_{Ki}p_j - U_{ij} - (U^*_{Ki}p_j - U^*_{ij})((U_K+1)/(U^*_K+1))$. The j-th row of $\hat{I}$ is $U_{KK}p_j - U_{jK} - (U^*_{KK}p_j - U^*_{jKj})((U_K+1)/(U^*_K+1))$.

The above matrix has the same rank as the matrix A. So, it is easy to see that the rank of A_i is K for some $i \in \{1,...,K-1\}$ if and only if an element of either the submatrix $\hat{X}$ or the submatrix $\hat{I}$ is different from 0. Hence, (U^*, f^*) is identified if and only if for some i and j $(1 \le i \le K; 1 \le j \le K-1)$, some p_j, and some x

$$\text{(iv)}\quad \frac{U_{Ki}(x)p_j - U_{ij}(x)}{U_K(x)+1} \neq \frac{U^*_{Ki}(x)p_j - U^*_{ij}(x)}{U^*_K(x)+1}.$$

It is easy to verify that this is satisfied if and only if for some x and some i, k, where $(1 \le i \le K; 1 \le k \le K-1)$,

$$\text{(v)}\quad \frac{U_{ij}(x)}{U_K(x)+1} \neq \frac{U^*_{ij}(x)}{U^*_K(x)+1}.$$

To see this, note if (v) is not satisfied, then (iv) can not be satisfied so that (iv) $\Rightarrow$ (v). To see that (v) implies (iv), divide (v) into the case in which (v) is satisfied when either i or k or both equal K and the case in which $\forall x$, $\forall i, k$ such that $i = K$ or $k = K$ or $i = k = K$, $U_{ik}(x)/(U_K(x)+1) = U^*_{ik}(x)/(U^*_K(x)+1)$. In the first case, it is always possible to find j and p_j such that (iv) is satisfied. In the second case, (v) implies that $\exists i, j\ (< K)$ and x such that $U_{ij}(x)/(U_K(x)+1) \neq U^*_{ij}(x)/(U^*_K(x)+1)$, and the definition of the case implies that $U_{Ki}(x)/(U_K(x)+1) = U^*_{Ki}(x)/(U^*_K(x)+1)$; so, (iv) is satisfied.

This completes the proof.

PROOF OF COROLLARY: It suffices to show that if $U \in W$ and $U \neq U^*$, there exists some x and some i and j, where $(1 \le i \le K; 1 \le j \le K-1)$, such that

$$\text{(vi)}\quad \frac{U_{ij}(x)}{U_K(x)+1} \neq \frac{U^*_{ij}(x)}{U^*_K(x)+1}.$$

Suppose then that $U \in W$ and $U \neq U^*$. Then, since $U(\bar{x}) = U^*(\bar{x})$, it must be that for so x and i, $U_i(x) \neq U^*_i(x)$. But then, since $DU(\bar{x}) = DU^*(\bar{x})$, it must be that for some x and j, $U_{ij}(x) \neq U^*(x)$. Since $U_K(x) = U^*_K(x) = g(x)$, it follows that $i \neq K$, and hence, (vi) is satisfied.

References

BROWN, B.W. (1983) "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 51, 175–196.

BROWN, B.W. and M.B. WALKER (1989) "The Random Utility Hypothesis and Inference in Demand Systems," *Econometrica*, 57, 815–829.

HAHN, J. (1994) "A Consistent Semiparametric Estimation of the Average Consumer Surpluses," mimeo, University of Pennsylvania.

HOUTHAKKER, H.S. (1950) "Revealed Preference and the Utility Function," *Economica NS* 17, 159–174.

HUBER, P.J. (1981) *Robust Statistics.* John Wiley and Sons.

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

McELROY, M.B. (1981) "Duality and the Error Structure in Demand Systems," Discussion Paper #81–82, Economics Research Center/NORC.

McELROY, M.B. (1987) "Additive General Error Models for Production, Cost, and Derived Demand or Share Systems," *Journal of Political Economy*, 95, 737–757.

McFADDEN, D. (1974) "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka (ed.), *Frontiers in Econometrics.* New York: Academic.

MANSKI, C.F. (1983) "Closest Empirical Distribution Estimation," *Econometrica*, 51(2), 305–320.

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

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

NEWEY, W.K. and J.L. POWELL (1989) "Instrumental Variables Estimation for Nonparametric Models," mimeo.

RAO, R. (1962) "Relationships Between Weak Convergence and Uniform Convergence of Measures with Applications," *Annals of Mathematical Statistics*, 33, 659–681.

ROEHRIG, C.S. (1988) "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56, 433–447.

UZAWA, H. (1971) "Preference and Rational Choice in the Theory of Consumption." Chapter 1 in J. Chipman, L. Hurwicz, M. Richter, and H. Sonnenschein (eds), *Preference, Utility, and Demand.* New York: Harcourt, Bruce, and Jovanovich.