HETEROGENEOUS CHOICE

Rosa L. Matzkin*

Department of Economics

Northwestern University

January 2007

This paper was presented as an invited lecture at the World Congress of the Econometric Society, London, August 2005. I am grateful to the National Science Foundation for financial support through grants SES-0551272, BCS 0433990, and SES 0241858. I have benefitted from discussions with and comments of Richard Blundell, Donald Brown, James Heckman, Arthur Lewbel, Daniel McFadden, Whitney Newey, Viktor Soubbotine, participants at the World Congress of the Econometric Society, and participants at the Harvard/MIT econometrics workshop and the Yale University Cowles Lunch series, where parts of this paper were presented.

1. Introduction

This paper considers nonparametric identification in models with nonadditive unobservables, where choices are either the dependent or the explanatory variables. For models where choices are dependent variables, the paper presents some of the latest developments in nonparametric identification, in the context of two popular models of consumer demand: the classical consumer demand and discrete choice. For models where choices are explanatory variables, it discusses several of the methods that have recently been developed to deal with their endogeneity.

Choice is the selection of an alternative from a set of alternatives. In many economic models, the role of economic agents is to make choices. Workers choose how to allocate their time between leisure and work-time. Consumers choose how to allocate their income between consumption and savings. Firms choose what and how much to produce. Governments choose taxes and subsidies. The usual assumption made in these economic models is that when making a choice, each economic agent has an objective function, and the choice is made so as to maximize that objective function over the set of possible alternatives.

Observable choices, made either by an agent or by different agents, are typically heterogeneous. This heterogeneity may be the result of heterogeneous choice sets, or of heterogenous objective functions. Some of this heterogeneity may be explained by different values of observable variables. Some may be explained by different values of unobservable variables. Understanding the source and shape of heterogeneity in choices is important, among other things, for accurately predicting behavior under new environments, and for evaluating the differential effect of policies over heterogeneous populations.

A usual approach in many econometric models has proceeded in the past by using economic theory to derive a relationship only between observable variables. Unobservable variables would then be added to the relationship, as an after-thought. An alternative approach proceeded by explicitly deriving the relationship between the underlying unobservable variables and the dependent variables. In these models, variables such as unobservable attributes of alternatives, unobservable taste for work, and unobservable productivity shocks were explicitly incorporated into the relationship fitted to the data. This typically resulted in models where the unobservable variables entered into the relationship in nonlinear, nonadditive ways (Heckman (1974), McFadden (1974), Heckman and Willis (1977), Lancaster (1979)). While those previous studies were parametric, in the sense that the underlying functions in the model as well as the distribution of the unobservable variables were specified up to a finite dimensional parameter, a large literature has evolved since then, following Manski (1975) and Heckman and Singer (1984), which relaxed those parametric assumptions. The present state of the literature allows us to consider models in which we can incorporate multidimensional, nonadditive unobservables in nonparametric models. (See, for example, the recent survey paper by Matzkin (2006), as well as the recent papers by Chesher (2006) and Imbens (2006).)

The outline of the paper is as follows. In the next section, we deal with the classical consumer demand model. We first consider a model with a distribution of consumers, each choosing a bundle of commodities from a linear budget set. We present conditions under which the distribution of preferences is identified, and show how one can recover the utility functions and the distribution of tastes from the distribution of demand. We then show how a large support condition on observable variables, which generate heterogeneity of tastes, can substitute for the large support condition on prices required for our first set of results. Nonlinear budget sets and situations where one observes variation in income but not in prices are also analyzed. Section 3 deals with discrete choice models. Section 4 deals with some of the latest methods that have been developed to deal with situations where explanatory variables are endogenous, in nonparametric models with nonadditive unobservables. It describes how that literature can be applied to models with heterogeneous choice. Section 5 concludes.

2. Classical consumer demand

Consumer behavior is arguably one of the most important element that are taken into account when considering policies and when determining actions by firms. The analysis of changes in income tax, interest rates, the proper design of cost of living indices, the pricing decision by firms, and the evaluation of changes in the welfare of consumers due to government policies require a good understanding of consumer behavior.

The classical model of consumer behavior is typically described as one where a consumer faces either a linear or a nonlinear budget set, which is determined by prices and the consumer's income. The consumer has preferences over commodities, and chooses the bundle of commodities that maximizes his preferences. Since consumers' preference might be different across individuals, one would like these preferences to depend on observable and on unobservable characteristics. The consumer's optimization problem can then be described as

$$(2.1) \qquad \begin{array}{ll} Max_y & V(y,z,\varepsilon) \\ s.t. & y \in B(w) \end{array}$$

where $V(\cdot, z, \varepsilon)$ is a utility function representing the preferences of the consumer over bundles of commodities, y; $z \in R^L$ denotes a vector of observable socioeconomic characteristics of the individual, $\varepsilon \in R^S$ denotes a vector of unobservable characteristics of the individual, and $B(w)$ denotes a subset of R^K, indexed by the observable vector w. In a standard example, w would be the vector of prices of the products and income of the consumer, z may denote age, profession, and/or level of education, and ε may denote a vector representing the different tastes of the consumer for the various products. The solution to this optimization problem, if unique, will be a vector $y = d(z, w, \varepsilon)$, such that $y \in B(w)$ and

$$V(d(z,w,\varepsilon), z, \varepsilon) > V(y', z, \varepsilon)$$

for all $y' \in B(w)$ such that $y' \neq y$.

The observable variables are w, z and the choice y. The questions that one may be interested in analyzing are (i) what restrictions the optimization assumption places on the distribution of y given (z, w), (ii) given the distribution of (y, z, w), under what conditions we can recover the function V and the distribution of ε, and (iii) suppose that the conditions for identification of V and the distribution of ε are not satisfied, then, what features of V and the distribution of ε can we recover.

Suppose that, in model (2.1), preferences did not depend on the values of unobservable variables, ε. Assume that the utility function V and the budget set $B(w)$ are such that the solution to the optimization problem is always unique. Then, on each budget set, $B(w)$, and for each z, one would observe only one choice, $d(z, w)$. In this case, the distribution of y given (z, w) would be degenerate. If the budget set were linear, then, for each z, $d(z, w)$ would satisfy the Slutsky conditions. And from $d(z, w)$, one could recover the utility function $V(y, z)$ using integrability or revealed preference theory. In this deterministic model, one could make use of the extensive literature on revealed preference, integrability, and duality theories (Samuelson (1938), Houtthaker (1950), Richter (1966), Afriat (1967), McFadden (1978), MasColell (1977, 1978), Epstein (1981), Yatchew (1985), Matzkin (1991b), and Matzkin and Richter (1991), among many others). These allow one to study characterizations of optimization and to infer the preferences generating the observable demand function of an individual.

When, on the other side, preferences depend on the value of unobservable variables, ε, then, on each observable budget set, $B(w)$, and for each z, one observes a distribution of choices, y, which are generated by the distribution of ε. A common approach in this situation proceeds by deriving the conditional expectation of the distribution of demand, and applying the theoretical results on consumer demand to this conditional expectation, as if it represented the deterministic demand of a representative consumer. On any budget set, any observable choice that is different from the conditional expectation of the distribution of demand on that budget set is represented by an unobservable random term that is added to that conditional expectation. However, in many situations, such a procedure might not be desirable. First, only under very strong assumptions on the preferences of all consumers one could guarantee that such representative consumer acted as if maximizing some utility function. Moreover, as pointed out and shown in McElroy (1981, 1987), Brown and Walker (1989, 1995), and Lewbel (2001), the additive unobservables tacked on to the deterministic relationship will typically be functionally dependent on the variables, w, that determine the choice set. One could alleviate these situations using parametric specifications for the utility V and/or the distribution of unobservable random terms, as suggested in those papers and their references (see, for example, Barten (1968)). We opt, instead, for a substantially more general, nonparametric analysis, that deals explicitly with unobservable taste heterogeneity.

2.1. Linear choice sets, price and income variation

The most common model of classical consumer demand is where the budget set is linear. In these models, w represents an observable vector, p, of K prices, and an observable variable, I, which denotes income. The choice (budget) set of the consumer is

$$B(p, I) \;=\; \left\{y \in R_{+}^{K} \mid p \cdot y \leq I\right\}$$

For each vector of observable socioeconomic characteristics, z, and unobservable tastes, ε, the preferences of the consumer over bundles, y, are represented by a function, $V(\cdot, z, \varepsilon)$, which is strictly increasing and strictly quasiconcave in y. For simplicity, we will strengthen this to consider only preferences that admit a strictly concave, twice continuously differentiable utility function. In addition, to avoid dealing with nonnegativity constraints, we will allow consumption to be negative. Denote the choice of a consumer with utility function $V(y, z, \varepsilon)$, facing prices p and income I, by

$$y = d\,(p, I, z, \varepsilon)$$

Then,

$$d\,(p, I, z, \varepsilon) = \arg\max_{\widetilde{y}} \{V(\widetilde{y}, z, \varepsilon) \mid \ p \cdot \widetilde{y} \leq I\}$$

For notational simplicity we will eliminate for the moment the dependence of V on z. Moreover, for the cases that we want to deal with later on, it will be important that the dimension of the vector ε be equal to the number of commodities minus 1. Hence, we will assume that the utility function is

$$V(y_1, ..., y_K, \varepsilon_1, ..., \varepsilon_{K-1})$$

The homogeneity of the budget constraint implies that we can normalize the value of one price. Hence, we set $p_K = 1$. The strict monotonicity of V in y guarantees that at any solution to the maximization problem, the budget constraint will be binding. Hence, we can restrict attention to those commodity bundles that satisfy the budget constraint. This allows us to substitute y_K for $y_K = I - \sum_{k=1}^{K-1} y_k p_k$, and solve for the unconstrained maximizer of

$$V\left(y_1, ..., y_{K-1}, I - \sum_{k=1}^{K-1} y_k p_k, \varepsilon_1, ..., \varepsilon_{K-1}\right)$$

The first order conditions for such maximization are, for $k = 1, ..., K-1$,

$$V_k\left(y_1, ..., y_{K-1}, I - \sum_{k=1}^{K-1} y_k p_k, \varepsilon_1, ..., \varepsilon_{K-1}\right) - V_K\left(y_1, ..., y_{K-1}, I - \sum_{k=1}^{K-1} y_k p_k, \varepsilon_1, ..., \varepsilon_{K-1}\right) p_k = 0$$

The second order conditions are satisfied, for each $(\varepsilon_1, ..., \varepsilon_{K-1})$, by the strict concavity of the utility function V. The first order conditions represent a system of $K-1$ simultaneous equations, with $K-1$ endogenous variables, $(y_1, ..., y_{K-1})$, and K exogenous variables, $p_1, ..., p_{K-1}, I$. The reduced form system can be expressed as a system of demand functions:

$$\begin{aligned} y_1 &= d_1(p_1, ..., p_{K-1}, I, \varepsilon_1, ..., \varepsilon_{K-1}) \\ y_2 &= d_2(p_1, ..., p_{K-1}, I, \varepsilon_1, ..., \varepsilon_{K-1}) \\ &\quad \cdot \quad \cdot \quad \cdot \\ y_{K-1} &= d_{K-1}(p_1, ..., p_{K-1}, I, \varepsilon_1, ..., \varepsilon_{K-1}) \end{aligned}$$

Since V is not specified parametrically, each function in this system is a nonparametric function, which possesses $K-1$ unobservable arguments. On each budget set, characterized by a vector of prices and income (p, I), the distribution of demand is generated by the distribution of $(\varepsilon_1, ..., \varepsilon_{K-1})$.

2.1.1. Invertibility of demand functions in the unobservables

A condition that, as we will show, will allow us to deal with a system of equations where each function depends on a vector of unobservable variables, is invertibility of the demand functions in that vector of unobservable variables. Invertibility guarantees that on each budget set, each observable choice corresponds to one and only one vector of the unobservable variables. It allows us to express the system of demand functions in terms of functions, $r_1, .., r_{K-1}$, such that

$$\begin{aligned} \varepsilon_1 &= r_1(y_1, ..., y_{K-1}, p_1, ..., p_{K-1}, I) \\ \varepsilon_2 &= r_2(y_1, ..., y_{K-1}, p_1, ..., p_{K-1}, I) \\ &\quad \cdot \quad \cdot \quad \cdot \\ \varepsilon_{K-1} &= r_{K-1}(y_1, ..., y_{K-1}, p_1, ..., p_{K-1}, I) \end{aligned}$$

Under invertibility, this system is obtained by solving for ε the system of demand equations.

Invertibility is not an innocuous assumption. When $K = 2$, it is satisfied if the demand function is either strictly increasing or strictly decreasing in ε. The following example is considered in Matzkin (2003):

Example 1: *Suppose that the consumer's problem is to maximize the random utility function*

$$v(y_1, y_2) + w(y_1, \varepsilon)$$
$$\textit{subject to} \quad p\, y_1 + y_2 \leq I$$

Assume that the functions v *and* w *are twice continuously differentiable, strictly increasing and strictly concave, and that* $\partial^2 w(y_1, \varepsilon)/\partial y_1 \partial \varepsilon > 0$. *Then, the demand function for* y_1 *is invertible in* ε

To see this, note that the value of y_1 that solves the maximization of

$$v\,(y_1, I - py_1) + w(y_1, \varepsilon)$$

satisfies

$$v_1\,(y_1, I - py_1) - v_2\,(y_1, I - py_1)\;\; p + w_1\,(y_1, \varepsilon) = 0$$

and

$$v_{11}\,(y_1, I - py_1) - 2\; v_{12}\,(y_1, I - py_1)\, p + v_{22}\,(y_1, I - py_1)\;\; p^2 + w_{11}\,(y_1, \varepsilon) < 0$$

where v_i, v_{ij}, w_i, and w_{ij} denote the partial derivatives of v, v_i, w, and w_i with respect to the $i-th$, $j-th$, $i-th$, and $j-th$ coordinate. By the Implicit Function Theorem, $y_1 = d\,(p, I, \varepsilon)$ that solves the first order conditions exists and satisfies for all p, I, ε,

$$\begin{aligned} & \frac{\partial d\,(p, I, \varepsilon)}{\partial \varepsilon} \\ = \;& -\frac{w_{12}\,(y_1, \varepsilon)}{v_{11}\,(y_1, I - py_1) - 2\; v_{12}\,(y_1, I - py_1)\, p + v_{22}\,(y_1, I - py_1)\;\; p^2 + w_{11}\,(y_1, \varepsilon)} \\ > \;& 0 \end{aligned}$$

Hence, the demand function for y_1 is invertible in ε.■

Example 1 demonstrates that some utility functions can generate demand functions that are strictly increasing in ε. This example also suggests that the demand functions generated from many other utility functions will not necessarily satisfy this condition.

When $K > 2$, invertibility requires even stronger conditions. Brown and Matzkin (1998) derived invertible demand functions by specifying a utility function of the form

$$V(y_1, ..., y_K, \varepsilon_1, ..., \varepsilon_{K-1}) = U(y_1, ..., y_K) + \sum_{k=1}^{K-1} \varepsilon_k \ y_k + y_K$$

where U is twice differentiable and strictly concave. In this case, the first order conditions for the maximization of V subject to the constraint that $y_K = I - \sum_{k=1}^{K-1} y_k p_k$, are

$$\begin{aligned}
\varepsilon_1 &= \left(U_K\left(y_1, ...y_{K-1}, I - \sum_{k=1}^{K-1} y_k p_k\right) + 1\right) p_1 - U_1\left(y_1, ...y_{K-1}, I - \sum_{k=1}^{K-1} y_k p_k\right) \\
\varepsilon_2 &= \left(U_K\left(y_1, ...y_{K-1}, I - \sum_{k=1}^{K-1} y_k p_k\right) + 1\right) p_2 - U_2\left(y_1, ...y_{K-1}, I - \sum_{k=1}^{K-1} y_k p_k\right) \\
& \quad \cdot \quad \cdot \quad \cdot \\
\varepsilon_{K-1} &= \left(U_K\left(y_1, ...y_{K-1}, I - \sum_{k=1}^{K-1} y_k p_k\right) + 1\right) p_{K-1} - U_{K-1}\left(y_1, ...y_{K-1}, I - \sum_{k=1}^{K-1} y_k p_k\right)
\end{aligned}$$

Hence, for each vector $(y_1, ..., y_{K-1})$, there exists only one value for $(\varepsilon_1, ..., \varepsilon_{K-1})$ satisfying the first order conditions.

Brown and Calsamiglia (2003), considering tests for random utility maximization, specified a slightly more restrictive utility function, which differs from the one in Brown and Matzkin (1998) mainly in that the twice differentiable and strictly concave function U depends only on $(y_1, ..., y_{K-1})$, and not on y_K :

$$V(y_1, ..., y_K, \varepsilon_1, ..., \varepsilon_{K-1}) = U(y_1, ..., y_{K-1}) + \sum_{k=1}^{K-1} \varepsilon_k \ y_k + y_K$$

The first order conditions for the maximization of U subject to the constraint that $y_K = I - \sum_{k=1}^{K-1} y_k p_k$, in this case have the much more simplified form:

$$\begin{aligned}
\varepsilon_1 &= p_1 - U_1(y_1, ...y_{K-1}) \\
\varepsilon_2 &= p_2 - U_2(y_1, ...y_{K-1}) \\
& \quad . \quad . \quad . \\
\varepsilon_{K-1} &= p_{K-1} - U_{K-1}(y_1, ...y_{K-1})
\end{aligned}$$

The optimal quantities of $y_1, ..., y_{K-1}$ depend on prices, p, and unobservable tastes, $(\varepsilon_1, ..., \varepsilon_{K-1})$, only through the vector $(p_1 - \varepsilon_1, ..., p_{K-1} - \varepsilon_{K-1})$. The strict concavity of U imply that this system has a unique solution, which is of the form

$$\begin{aligned}
y_1 &= d_1(p_1 - \varepsilon_1, ..., p_{K-1} - \varepsilon_{K-1}) \\
y_2 &= d_2(p_1 - \varepsilon_1, ..., p_{K-1} - \varepsilon_{K-1}) \\
& \quad . \quad . \quad . \\
y_{K-1} &= d_{K-1}(p_1 - \varepsilon_1, ..., p_{K-1} - \varepsilon_{K-1})
\end{aligned}$$

Beckert and Blundell (2005) consider more general utility functions, whose demands are invertible in the unobservable variables.

A very important question in demand models is whether one can recover the distribution of tastes and the utility function U, from the distribution of demand. The answer is much easier to obtain when demand depends on one unobservable taste than when it depends on many unobservable tastes. The first case typically occurs when the number of commodities is two. The latter typically occurs when this number of larger than two.

2.1.2. Identification when $K = 2$

As Example 1 illustrates, in some circumstances, when the number of commodities is 2, it is possible to restrict the utility function to guarantee that the demand function for one of those commodities is strictly increasing in a unique unobservable taste variable, ε. In such a case, the analysis of nonparametric identification of the demand function and the distribution of ε can proceed using the methods developed in Matzkin (1999, 2003). The general situation considered in

Matzkin (1999) is one where the value of an observable dependent variable, Y, is determined by an observable vector of explanatory variables, X, and an unobservable variable, ε, both arguments in a nonparametric function $m(X, \varepsilon)$. It is assumed that m is strictly increasing in ε, over the support of ε, and ε is distributed independently of X. Under these conditions, the unknown function m and the unknown distribution, F_ε, of ε satisfy for every x, ε, the equation

$$F_\varepsilon(\varepsilon) = F_{Y|X=x}(m(y, \varepsilon))$$

where $F_{Y|X=x}$ denotes the conditional distribution of Y given $X = x$. (See Matzkin (1999).) In our particular case, we can let Y denote the observable demand for one of the two commodities, $X = (p, I)$ denote the vector of observable explanatory variables, and ε denote the unobservable taste. Hence, if the demand is strictly increasing in ε, it follows by analogy that, for all (p, I) and ε,

$$F_\varepsilon(\varepsilon) = F_{Y|p,I}(d(p, I, \varepsilon))$$

where F_ε is the cumulative distribution of ε and $F_{Y|p,I}$ is the cumulative distribution of Y given (p, I). It is clear from this functional equation that without any further restrictions, it is not possible to identify both, F_ε and d, nonparametrically.

2.1.2.1. **Normalizations**

One approach that can be used to deal with the non-identification of the unknown functions is to specify a normalization for one of them. For this, it is convenient to first determine, for any pair, (d, F_ε), the set of pairs, $\left(\widetilde{d}, F_{\widetilde{\varepsilon}}\right)$, which are observationally equivalent to (d, F_ε). We say that $\left(\widetilde{d}, F_{\widetilde{\varepsilon}}\right)$ is *observationally equivalent* to (d, F_ε) if the distribution of Y given (p, I) generated by $\left(\widetilde{d}, F_{\widetilde{\varepsilon}}\right)$ equals almost surely, in (p, I), the distribution of Y given (p, I) generated by (d, F_ε).

We assume, as above, that d and $\widetilde{d}$ are strictly increasing in ε and $\widetilde{\varepsilon}$, over their respective convex supports E and $\widetilde{E}$, and that ε and $\widetilde{\varepsilon}$ are distributed independently of X. Moreover, we will assume that F_ε and $F_{\widetilde{\varepsilon}}$ are continuous. We can characterize d and $\widetilde{d}$ by their inverse functions, v and $\widetilde{v}$, defined by $y = d(p, I, v(y, p, I))$ and $y = \widetilde{d}(p, I, \widetilde{v}(y, p, I))$. Then, by the analysis in Matzkin (2003), it follows that, under our assumptions, $\left(\widetilde{d}, F_{\widetilde{\varepsilon}}\right)$ is observationally equivalent to (d, F_ε) iff for some continuous and strictly increasing $g : E \to \widetilde{E}$,

$$\widetilde{v}(y,P,I) = g\left(v(y,p,I)\right) \quad \text{and} \quad F_{\widetilde{\varepsilon}}(e) = F_{\varepsilon}\left(g^{-1}(e)\right).$$

This implies that the function d is identified up to a strictly increasing transformation on ε. Or, in other words, that the function d can be identified only on the ordering of the values of ε. In particular, if on a given budget set, the demand for the commodity of one individual is different than the demand for the same commodity of another individual, then, one can identify which of the individuals has a larger value of ε, but one cannot identify the actual values of their corresponding tastes ε.

The implication of this analysis is that to select from each set of observationally equivalent pairs, a unique pair, one needs to impose restrictions on the set of functions v guaranteeing that no two functions can be expressed as strictly increasing transformation of each other. Alternatively, one may impose restrictions on the set of distributions guaranteeing that the random variables corresponding to any two distributions are not strictly increasing transformations of each other.

To normalize the set of inverse functions so that no two functions are strictly increasing transformations of each other, Matzkin (1999) proposed restricting all inverse functions v to satisfy that, at some value of the vector of explanatory variables, $\left(\overline{p},\overline{I}\right)$, and for all y,

$$v\left(y,\overline{p},\overline{I}\right) = y$$

This guarantees that if $\widetilde{v}$ is such that for some strictly increasing g and for all y,p,I

$$\widetilde{v}\left(y,p,I\right) = g\left(v\left(y,p,I\right)\right), \text{ and}$$

$$\widetilde{v}\left(y,\overline{p},\overline{I}\right) = y$$

then $\widetilde{v} = v$. The demand functions corresponding to inverse functions satisfying the above restriction satisfy for all $\widetilde{\varepsilon}$,

$$\widetilde{d}\left(\overline{p},\overline{I},\widetilde{\varepsilon}\right) = \widetilde{\varepsilon}$$

Hence, as long as one restricts the set of demand functions $\widetilde{d}$ to be such that at one budget $\left(\overline{p},\overline{I}\right)$, for all values of $\widetilde{\varepsilon}$

$$\widetilde{d}\left(\overline{p},\overline{I},\widetilde{\varepsilon}\right)=\widetilde{\varepsilon}$$

it will not be possible to find two demand functions, d and $\widetilde{d}$ and two distributions F_{ε} and $F_{\widetilde{\varepsilon}}$, such that (d,F_{ε}) and $\left(\widetilde{d},F_{\widetilde{\varepsilon}}\right)$ are observationally equivalent. This normalization can be seen as a generalization of a linear demand with additive unobservables. In the latter, when the value of the explanatory variables equals zero, demand equals the value of the unobservable random term, after restricting the intercept of the linear function to be zero. (Note that we are not restricting the expectation of ε to be zero, as it is typically done in linear models. Hence assuming a zero intercept is not more restrictive than the standard assumption made in linear models.)

Under the above normalization, the distribution of ε can be read off from the distribution of the demand when $(p,I)=\left(\overline{p},\overline{I}\right)$. Specifically,

$$F_{\varepsilon}(e)=F_{Y|(p,I)=\left(\overline{p},\overline{I}\right)}(e).$$

Then, for any $\widetilde{p},\widetilde{I},\widetilde{e}$

$$d\left(\widetilde{p},\widetilde{I},\widetilde{e}\right)=F^{-1}_{Y|(p,I)=\left(\widetilde{p},\widetilde{I}\right)}\left(F_{Y|(p,I)=\left(\overline{p},\overline{I}\right)}(\widetilde{e})\right)$$

Alternatively, one can specify the distribution of ε. In such case,

$$d\left(\widetilde{p},\widetilde{I},\widetilde{e}\right)=F^{-1}_{Y|(p,I)=\left(\widetilde{p},\widetilde{I}\right)}\left(F_{\varepsilon}(\widetilde{e})\right).$$

In particular, when F_{ε} is the distribution of a $U(0,1)$ random variable, the demand function is a conditional quantile function, whose estimation has been extensively studied. (See Imbens and Newey (2001) for the use of this normalization, and Koenker (2005) for estimation methods for conditional quantiles.)

2.1.2.2. **Restrictions**

In some cases, economic theory might imply some properties on either the unknown function, d, or the unknown distribution, F_{ε}. These restrictions might allow us to identify these functions with fewer normalizations. Suppose for example that $V(y_1,y_2,\varepsilon)=\widetilde{V}\left(y_1,y_2+\varepsilon\right)$ for some function

$\widetilde{V}$. The demand function for $Y = y_1$ derived from such a utility function can be shown to have the form $d(p, I+\varepsilon)$, when p is the price of one unit of y_1 and the price of one unit of y_2 is normalized to 1. In this case, we can identify the demand function d and the distribution of ε by only normalizing the value of the demand function at one point, e.g. $d(\overline{p}, \overline{t}) = \overline{y}$. In such a case, F_ε can be recovered from the distribution of demand at different values of income

$$F_\varepsilon(\varepsilon) = F_{Y|p=\overline{p}, I=\overline{t}-\varepsilon}(\overline{y})$$

and then, for any $\left(\widetilde{p}, \widetilde{I}, \varepsilon\right)$

$$d\left(\widetilde{p}, \widetilde{I}+\varepsilon\right) = F^{-1}_{Y|(p,I)=\left(\widetilde{p},\widetilde{I}\right)}\left(F_{Y|\overline{p},I=\overline{t}-\varepsilon}(\overline{y})\right)$$

2.1.2.3. **Functionals of the demand**

Rather than considering different normalizations or restrictions, which would allow us to identify the unknown functions and distributions in the model, we might ask what can be identified without them. For example, given a consumer with unobservable taste ε, we might want to determine by how much his consumption would change if his budget set changed from $B(\widetilde{p}, \widetilde{I})$ to $B(p', I')$. This change in demand can be identified without normalizations or restrictions on the unknown functions, other than assuming that that ε is distributed independently of (p, I), the demand function is strictly increasing in ε, and F_ε is strictly increasing at ε. Specifically,

$$d(p', I', \varepsilon) - d\left(\widetilde{p}, \widetilde{I}, \varepsilon\right)$$

$$= \quad F^{-1}_{Y|(p,I)=(p',I')}\left(F_{Y|(p,I)=\left(\widetilde{p},\widetilde{I}\right)}(y_1)\right) - y_1$$

where y_1 is the observed consumption of the individual when his budget is (p, I). (See Matzkin (2006).) When to the above assumptions one adds differentiability, one can obtain, using arguments as in Matzkin (1999) and Chesher (2003) that

$$\frac{\partial d\left(\widetilde{p},\widetilde{I},\varepsilon\right)}{\partial\left(p,I\right)}=-\left[\frac{\partial F_{Y|(p,I)=\left(\widetilde{p},\widetilde{I}\right)}\left(d\left(\widetilde{p},\widetilde{I},\varepsilon\right)\right)}{\partial y}\right]^{-1}\frac{\partial F_{Y|(p,I)=\left(\widetilde{p},\widetilde{I}\right)}\left(d\left(\widetilde{p},\widetilde{I},\varepsilon\right)\right)}{\partial(p,I)}$$

(See also Altonji and Matzkin (1997) and Athey and Imbens (2006) for related expressions.)

2.1.3. Identification when $K > 2$

When the number of unobservable taste variables, $\varepsilon_1, ..., \varepsilon_{K-1}$, is larger than one, the demand function for each commodity typically depends on all these unobservable variables. For example, in individual consumer demand, the chosen quantity of entertainment will depend on the taste of the individual for housing, since spending a large amount in housing will typically decrease the amount spent in entertainment. If one could impose some type of separability, so that the demand function depended on a one dimensional function of the unobservables, then the methods for $K = 2$ might be used. When, however, this is not desirable, one needs to consider identification of functions whose arguments are unobservable. The analysis of identification in this case may be performed using results about identification of nonparametric simultaneous equations.

Roehrig (1988), following the approach in Brown (1983), established a rank condition for nonparametric simultaneous equations, which when satisfied, guarantees that the model is not identified. Benkard and Berry (2004) showed that when Roehrig's rank condition is not satisfied, the model might not be identified. Matzkin (2005) developed a different rank condition, which can be used to determine identification of simultaneous equation models, in general, and of system of demands, in particular. Other methods that have been recently developed to either identify or estimate equations in systems with simultaneity are Newey and Powell (1989, 2003), Brown and Matzkin (1998), Darolles, Florens and Renault (2002), Hall and Horowitz (2003), Ai and Chen (2003), Matzkin (2004), Chernozhukov, Imbens, and Newey (2004), and Chernozhukov and Hansen (2005).

We follow Matzkin (2005). Assume that the first order conditions for utility maximization subject to the budget constraint can be solved for ε, in terms of (y, p, I). Then, the model can be expressed as

$$\varepsilon = r\left(y, p, I\right)$$

We will assume that r is twice differentiable in (y, p, I), invertible in y, and the Jacobian determinant of r with respect to y, $|\partial r(y, p, I)/\partial y| > 0$. The later is the extension to multidimensional situations of the monotonicity requirement in the single equation, univariate unobservable demand function $d(p, I, \varepsilon)$ analyzed above for the case where $K = 2$. In the $K = 2$ case, it was only necessary to require that d be strictly increasing in ε.

To analyze identification in the $K > 2$ case, we can proceed as in the $K = 2$ case, by first characterizing the sets of pairs $(r, f_\varepsilon), (\widetilde{r}, f_{\widetilde{\varepsilon}})$ that are observationally equivalent among them. We assume that ε and $\widetilde{\varepsilon}$ are distributed independently of (p, I) with differentiable densities f_ε and $f_{\widetilde{\varepsilon}}$, that $\widetilde{r}$ is twice differentiable in (y, p, I), invertible in y, and with Jacobian determinant with respect to y, $|\partial\widetilde{r}(y, p, I)/\partial y| > 0$. We will also assume that (p, I) has a differentiable density and convex support. Under these conditions, we can say that the pairs (r, f_ε) and $(\widetilde{r}, f_{\widetilde{\varepsilon}})$ are *observationally equivalent* if for all y, p, I

$$f_\varepsilon\left(r\left(y, p, I\right)\right)\left|\frac{\partial r\left(y, p, I\right)}{\partial y}\right| = f_{\widetilde{\varepsilon}}\left(\widetilde{r}\left(y, p, I\right)\right)\left|\frac{\partial \widetilde{r}\left(y, p, I\right)}{\partial y}\right|$$

This is clear because, when this condition is satisfied, both pairs (r, f_ε) and $(\widetilde{r}, f_{\widetilde{\varepsilon}})$ generate the same conditional densities, $f_{y|(p,I)}$, of y given (p, I) Under our assumptions, the distribution of y given p, I, which is generated by (r, f_ε) is

$$f_{y|p,I}(y) = f_\varepsilon\left(r\left(y, p, I\right)\right)\left|\frac{\partial r\left(y, p, I\right)}{\partial y}\right|$$

and this equals the conditional distribution of $f_{y|p,I}$ generated by $(\widetilde{r}, f_{\widetilde{\varepsilon}})$.

Example 2: *Suppose that for a twice differentiable function* $g : E \to \widetilde{E}$, *with Jacobian determinant* $|\partial g(\varepsilon)/\partial\varepsilon| > 0$

$$\widetilde{r}\left(y, p, I\right) = g\left(r(y, p, I)\right) \qquad \text{and} \qquad \widetilde{\varepsilon} = g\left(\varepsilon\right)$$

where $E \subset R^{K-1}$ is the convex support of ε, and $E \subset R^{K-1}$, *then,* $(\widetilde{r}, f_{\widetilde{\varepsilon}})$ *is observationally equivalent to* (r, f_ε).

To see this, note that in this case

$$\widetilde{\varepsilon} = g\left(\varepsilon\right) = g(r(y,p,I)) = \widetilde{r}(y,p,I)$$

Using the relationship $\widetilde{\varepsilon} = g(\varepsilon)$, to derive the distribution of ε from that of $\widetilde{\varepsilon}$, and substituting $\widetilde{r}(y,p,I)$ for $g(\varepsilon)$, we get that

$$f_{\varepsilon}\left(\varepsilon\right) = f_{\widetilde{\varepsilon}}\left(g(\varepsilon)\right) \, \left|\frac{\partial g(\varepsilon)}{\partial \varepsilon}\right| = f_{\widetilde{\varepsilon}}\left(\widetilde{r}\left(y,p,I\right)\right) \, \left|\frac{\partial g(r(y,p,I))}{\partial \varepsilon}\right|$$

Using the relationship $\widetilde{r}\left(y,p,I\right) = g\left(r(y,p,I)\right)$, we also get that

$$\left|\frac{\partial \widetilde{r}\left(y,p,I\right)}{\partial y}\right| = \left|\frac{\partial g(r\left(y,p,I\right))}{\partial y}\right| = \left|\frac{\partial g(r\left(y,p,I\right))}{\partial \varepsilon}\right| \left|\frac{\partial r\left(y,p,I\right))}{\partial y}\right|$$

Hence,

$$f_{\varepsilon}\left(r\left(y,p,I\right)\right) \, \left|\frac{\partial r\left(y,p,I\right)}{\partial y}\right| = f_{\widetilde{\varepsilon}}\left(\widetilde{r}\left(y,p,I\right)\right) \, \left|\frac{\partial \widetilde{r}\left(y,p,I\right)}{\partial y}\right|$$

■

Example 2 shows that invertible transformations, g, of $r(y,p,I)$, satisfying $|\partial g(\varepsilon)/\partial\varepsilon| > 0$, together with the associated vector of unobservables $\widetilde{\varepsilon} = g(\varepsilon)$, generate pairs $(\widetilde{r}, f_{\widetilde{\varepsilon}})$ that are observationally equivalent to (r, f_{ε}). One may wonder, as Benkard and Berry (2004) did, whether invertible transformations that depend on the observable exogenous variables, (p, I), as well as on ε might also generate pairs, $(\widetilde{r}, f_{\widetilde{\varepsilon}})$, that are observationally equivalent to (r, f_{ε}). When $K = 2$, this is not possible, in general. Strictly increasing transformation of ε that also depends on $x = (p, I)$ will generate a random variable, $\widetilde{\varepsilon} = g(\varepsilon, x)$, which is not distributed independently of (p, I). (Example 3.3 in Matzkin (2005)) provides a proof of this. See also Matzkin (2003).) Hence, when the number of commodities is 2, one can guarantee identification of the true pair (f_{ε}, r) in a set of pairs $(f_{\widetilde{\varepsilon}}, \widetilde{r})$ by eliminating from that set all pairs $(f_{\widetilde{\varepsilon}}, \widetilde{r})$ that are obtained by strictly increasing transformations, g, of ε. When, however, $K > 2$, invertible transformations of ε that depend on (p, I) might generate pairs that are also observationally equivalent to (r, f_{ε}). For general simultaneous equations models, this claim was recently shown by Benkard and Berry (2004, 2006).

Example 3 (Benkard and Berry (2006)) *Consider a simultaneous equation model with* $\varepsilon \in R^2$ *and*

$$\varepsilon = r\left(y, x\right)$$

where $y \in R^2$ *is the vector of observable endogenous variables and* $x \in R$ *is a vector of observable exogenous variables. Suppose that* ε *is* $N(0, I)$ *and let*

$$\widetilde{\varepsilon} = g\left(\varepsilon, x\right) = \Gamma(x)'\varepsilon$$

where $\Gamma(x)$ *is an orthonormal matrix that is a smooth function of* x*. Then,* $\widetilde{\varepsilon}$ *is* $N\left(0, \Gamma(x)'\Gamma(x)\right) = N(0, I)$*. Let*

$$\widetilde{r}(y, x) = \Gamma(x)'r(y, x)$$

Then,

$$(r, f_{\varepsilon}) \text{ and } (\widetilde{r}, f_{\widetilde{\varepsilon}}) \text{ are observationally equivalent}$$

To see this, note that for all y, x

$$\begin{aligned}
f_{Y|X=x}(y) &= f_{\varepsilon}\left(r\left(y, x\right)\right) \left|\frac{\partial r\left(y, x\right)}{\partial y}\right| \\
&= (2\pi)^{-1} \exp\left(-\frac{(r(y, x)'r(y, x))}{2}\right) \left|\frac{\partial r\left(y, x\right)}{\partial y}\right| \\
&= (2\pi)^{-1} \exp\left(-\frac{(r(y, x)'\Gamma(x)'\Gamma(x)r(y, x))}{2}\right) |\Gamma(x)| \left|\frac{\partial r\left(y, x\right)}{\partial y}\right| \\
&= (2\pi)^{-1} \exp\left(-\frac{(\widetilde{r}(y, x)'\widetilde{r}(y, x))}{2}\right) \left|\frac{\partial \widetilde{r}\left(y, x\right)}{\partial y}\right|
\end{aligned}$$

$$= \quad f_{\widetilde{\varepsilon}}\left(\widetilde{r}\left(y,x\right)\right) \quad \left|\frac{\partial \widetilde{r}\left(y,x\right)}{\partial y}\right|$$

Hence, (f_{ε}, r) and $(f_{\widetilde{\varepsilon}}, \widetilde{r})$ are observationally equivalent.■

Conditions for identification in nonparametric simultaneous equations with nonadditive unobservables were provided by Matzkin (2005). To describe these results, let X denote the vector of observable exogenous variables. Let r^*, f_{ε^*} denote the true function and true distribution. Denote by P a set of functions to which r^* is assumed to belong, and denote by Γ the set of densities to which f_{ε^*} is assumed to belong. Assume that all functions $r \in P$ are such that (i) r is twice continuously differentiable, (ii) for all y, x, $|\partial r(y,x)/\partial y| > 0$, and (iii) for all $x, \widetilde{\varepsilon}$ there exists a unique $\widetilde{y}$ satisfying $\widetilde{\varepsilon} = \widetilde{r}(\widetilde{y}, x)$. Assume that the vector of observable exogenous variables, X, has a differentiable density, f_X, and convex support in R^K. Further, assume that for all $f_{\varepsilon} \in \Gamma$, f_{ε} is continuously differentiable and with convex support; and that for all $r \in P$ and all $f_{\varepsilon} \in \Gamma$, the distribution of Y given $X = x$, which is generated by r and f_{ε}, is continuously differentiable and has convex support in R^G. Then, by Matzkin (2005) it follows that

Theorem (Matzkin (2005)): *Under the above assumptions (r^*, f_{ε^*}) is identified within $(P \times \Gamma)$ if for all $r, \widetilde{r} \in P$ and all $f_{\varepsilon} \in \Gamma$, there exists a value of (y,x) at which the density of (y,x) is strictly positive and the rank of the matrix*

$$B\left(y,x;r,\widetilde{r},f_{\varepsilon}\right) = \begin{pmatrix} \left(\frac{\partial \widetilde{r}(y,x)}{\partial y}\right)' & \Delta_y\left(y,x;\partial r,\partial^2 r,\partial\widetilde{r},\partial^2\widetilde{r}\right) - \frac{\partial \log(f_{\varepsilon}(r(y,x)))}{\partial \varepsilon}\frac{\partial r(y,x)}{\partial y} \\ \\ \left(\frac{\partial \widetilde{r}(y,x)}{\partial x}\right)' & \Delta_x\left(y,x;\partial r,\partial^2 r,\partial\widetilde{r},\partial^2\widetilde{r}\right) - \frac{\partial \log(f_{\varepsilon}(r(y,x)))}{\partial \varepsilon}\frac{\partial r(y,x)}{\partial x} \\ \end{pmatrix}$$

is strictly larger than G, where

$$\Delta_y\left(y,x;\partial r,\partial^2 r,\partial\widetilde{r},\partial^2\widetilde{r}\right) = \frac{\partial}{\partial y}\log\left|\frac{\partial r(y,x)}{\partial y}\right| - \frac{\partial}{\partial y}\log\left|\frac{\partial\widetilde{r}(y,x)}{\partial y}\right|$$

$$\Delta_x\left(y,x;\partial r,\partial^2 r,\partial\widetilde{r},\partial^2\widetilde{r}\right) = \frac{\partial}{\partial x}\log\left|\frac{\partial r(y,x)}{\partial y}\right| - \frac{\partial}{\partial x}\log\left|\frac{\partial\widetilde{r}(y,x)}{\partial y}\right| \blacksquare$$

The following example applies this result to a random utility model where the utility function is of the type considered in Brown and Calsamiglia (2003).

Example 4: *Suppose that the utility for a consumer with tastes ε is*

$$V(y_1,...,y_K,\varepsilon_1,...,\varepsilon_{K-1}) = U\left(y_1,...,y_{K-1}\right) + \sum_{k=1}^{K-1}\varepsilon_k\, y_k + y_K.$$

Denote the gradient of U by DU and the Hessian of U by D^2U. Let $\overline{y}$ denote a value of y. For fixed $\alpha \in R^{K-1}$ and $\Lambda \in R$, let W denote the set of functions U such that $DU(\overline{y}) = \alpha$ and $\left|D^2U(\overline{y})\right| = \Lambda$. Let ε^* denote a value of ε. Let Γ denote the set of densities, f_ε, such that (i) f_ε is differentiable, (ii) $f_\varepsilon(\varepsilon) > 0$ on a neighborhood of radius δ around ε^*, (iii) for all ε in the support of f_ε, $\partial\log(f_\varepsilon(\varepsilon))\partial\varepsilon = 0$ iff $\varepsilon = \varepsilon^*$, (iv) for all k, there exist two distinct values, ε' and ε'', in the $\delta-$neighborhood of ε^* such that $f_\varepsilon(\varepsilon')$, $f_\varepsilon(\varepsilon'') > 0$, $\partial f_\varepsilon(\varepsilon')/\partial\varepsilon_k$, $\partial f_\varepsilon(\varepsilon'')/\partial\varepsilon_k \neq 0$, $\partial\log(f_\varepsilon(\varepsilon'))\partial\varepsilon_k \neq \partial\log(f_\varepsilon(\varepsilon'')/\partial\varepsilon_k$, and for $j \neq k$, $\partial\log(f_\varepsilon(\varepsilon'))\partial\varepsilon_j = \partial\log(f_\varepsilon(\varepsilon'')/\partial\varepsilon_j = 0$.

Let W and the support of p be such for all y, for all $U \in W$, there exist a set of prices, Q, such that the density of p is uniformly bounded away from zero on Q and the range of $DU(y) - p$, when considered as a function of p over Q, is the δ neighborhood of ε^*. Then, if $U, \widetilde{U}$ belong to W and $D\widetilde{U} \neq DU$, there exist, for all $f_\varepsilon \in \Gamma$, values y, p such that the rank of the matrix $B\left(y,p;DU,D\widetilde{U},f_\varepsilon\right)$ is larger than $K-1$.

The proof of this result can be obtained by modifying an analogous theorem in Brown and

Matzkin (1998).

2.1.4. Recovering the distribution of utility functions when $K > 2$

An implication of the last example is that, if the observable price vector p has an appropriate support and the utility function is given by

$$(2.2) \qquad V(y_1,...,y_K,\varepsilon_1,...,\varepsilon_{K-1}) = U(y_1,...,y_{K-1}) + \sum_{k=1}^{K-1} y_k \cdot \varepsilon_k + y_K$$

then, the rank conditions for nonparametric identification of U and f_ε are satisfied. Under those conditions, we can constructively recover $DU(y)$ and f_ε from the distribution of demand, by applying the results in Section 6 of Matzkin (2005). (See Matzkin (2006).)

Example 4 (continued)**:** *Suppose that given any budget set* (p, I), *each consumer with unobservable tastes* $(\varepsilon_1,...,\varepsilon_{K-1})$ *chooses his demand by maximizing* $V(y_1,...,y_K,\varepsilon_1,...,\varepsilon_{K-1})$, as in (2), *subject to* $p \cdot y \leq I$. *Let* $f_{Y|p,I}$ *denote the density of demand generated by the density of* ε, *over the budget set determined by* (p, I). *Suppose that* U *belongs to the set* W, *and* f_ε *belongs to the set of densities* Γ. *Then, the density of* ε *and the gradient,* $DU(y)$, *of* U, *can be recovered from* $f_{Y|p,I}$, *as long as this conditional density is strictly positive at the values from which* DU *and* f_ε *are recovered. In particular, for* $(t_1,...,t_{K-1})$,

$$f_\varepsilon(t_1,...,t_{K-1}) = f_{Y|p_1=t_1+\alpha_1,...,p_{K-1}=t_{K-1}+\alpha_{K-1},I}(\overline{y}) \quad \Lambda^{-1}$$

and for y,

$$DU(y) = p^* - \varepsilon^*$$

where p^* *is the value of* p *that satisfies*

$$\frac{\partial f_{Y|p=p^*,I}(y)}{\partial p} = 0$$

2.2. Linear choice sets, income variation, no price variation

A critical requirement in the identification results in the previous section was that, given any bundle of commodities and any value for the marginal utility, one could always observe a price vector with the value necessary to identify that marginal utility. That requires a large support condition on these variables. In most situations, however, that large support condition is not satisfied. In fact, it is often the case that the available data corresponds to only a few values of prices. Blundell, Browning, and Crawford (2003) deal with this situation by ingeniously making use of observations on income. They note that, although the data corresponds to only a few price vectors, income across different individuals has a large support. They use this to estimate the income expansion paths of the average consumer. Assuming that the average consumer behaves as if maximizing a utility function, they use these choices to determine bounds on the demand of this average consumer on previously unobserved budget sets. These bounds are determined from reveled preference. (See also Blundell, Chen and Kristensen (2003), Blundell, Browning, and Crawford (2004), and Blundell, Browning, Crawford (2005).)

2.2.1. Bounds on the derivative of the distribution of demand with respect to price

The assumption that the average consumer behaves as if maximizing a utility function might be too restrictive in sone circumstances. Blundell and Matzkin (2005) and Blundell, Kristensen, and Matzkin (2005) analyze what can be said when it is only assumed that each consumer acts as if maximizing a utility. No specific restrictions are imposed on the average consumer. In other words, they consider a population of consumers such that each consumer possesses a utility function

$$V(y, \varepsilon)$$

On each budget set, characterized by a price vector, p, and an income I, the distribution of ε generates a distribution of demand over the budget set determined by p and I. This gives rise, as in the previous sections, to a system of demand functions

$$
\begin{aligned}
y_1 &= d_1\left(p, I, \varepsilon_1, ..., \varepsilon_{K-1}\right) \\
y_2 &= d_2\left(p, I, \varepsilon_1, ..., \varepsilon_{K-1}\right) \\
&\quad \cdot \;\cdot \;\cdot \\
y_{K-1} &= d_{K-1}\left(p, I, \varepsilon_1, ..., \varepsilon_{K-1}\right),
\end{aligned}
$$

which, for each (p, I) and $(\varepsilon_1, ..., \varepsilon_{K-1})$, solve the maximization of $V(y, \varepsilon)$ subject to the budget constraint. Assuming that the system is invertible in $(\varepsilon_1, ..., \varepsilon_{K-1})$, we can express it as

$$
\begin{aligned}
\varepsilon_1 &= r_1\left(p, I, y_1, ..., y_{K-1}\right) \\
\varepsilon_2 &= r_2\left(p, I, y_1, ..., y_{K-1}\right) \\
&\quad \cdot \;\cdot \;\cdot \\
\varepsilon_{K-1} &= r_{K-1}\left(p, I, y_1, ..., y_{K-1}\right)
\end{aligned}
$$

Making use of the variation in income, together with the implications of the Slutsky conditions for optimization, Blundell and Matzkin (2006) obtain bounds on the unobservable derivative of the density of demand with respect to price.

To provide an example of the type of restrictions that one can obtain, consider again the $K = 2$ case. Let $d(p, I, \varepsilon)$ denote the demand function for one of the commodities. If this demand function is generated from the maximization of a twice differentiable, strictly quasiconcave utility function, it satisfies the Slutsky conditions

$$
\frac{\partial d(p, I, \varepsilon)}{\partial p} + d(p, I, \varepsilon) \, \frac{\partial d(p, I, \varepsilon)}{\partial I} \leq 0
$$

Under the assumptions in Section 2.1.2.3, the derivative of the demand with respect to income, of a consumer that is observed demanding $y = d\left(p, I, \varepsilon\right)$ can be recovered from the distribution of demand over budget sets with a common price vector, p, and different income levels, I, by

$$
\frac{\partial d(p, I, \varepsilon)}{\partial I} = \left(\frac{\partial F_{Y|(p,I)}(y)}{\partial y} \right)^{-1} \left(\frac{\partial F_{Y|(p,I)}(y)}{\partial I} \right)
$$

Hence, an upper bound on the derivative of the demand with respect to price of a consumer that demands y at prices p and income I is given by

$$\frac{\partial d(p,I,\varepsilon)}{\partial p} \leq -\ y\ \left(\frac{\partial F_{Y|(p,I)}(y)}{\partial y}\right)^{-1}\left(\frac{\partial F_{Y|(p,I)}(y)}{\partial I}\right)$$

Note that ε need not be known to obtain such an expression.

2.2.1. Representation of demand

In some cases, one may be able to analyze demand by using a representation of the system of demand. Blundell, Kristensen, and Matzkin (2005) consider one such representation. Suppose, as above, that the system of demand functions is given by

$$\begin{aligned} y_1 &= d_1\left(p, I, \varepsilon_1, ..., \varepsilon_{K-1}\right) \\ y_2 &= d_2\left(p, I, \varepsilon_1, ..., \varepsilon_{K-1}\right) \\ &\quad . \ . \ . \\ y_{K-1} &= d_{K-1}\left(p, I, \varepsilon_1, ..., \varepsilon_{K-1}\right), \end{aligned}$$

A representation of this system, in terms of a different vector of unobservable variables, $\left(\eta_1, ..., \eta_{K-1}\right)$, can be obtained by defining these unobservables sequentially as

$$\begin{aligned} \eta_1 &= F_{Y_1|p,I}(y_1) \\ \eta_2 &= F_{Y_2|\eta_1,p,I}(y_2) \\ &\quad . \ . \ . \\ \eta_{K-1} &= F_{Y_{K-1}|\eta_1,\eta_2,...,\eta_{K-1},p,I}(y_{K-1}) \end{aligned}$$

where $F_{Y_1|p,I}$ is the conditional distribution of Y_1 given (p, I), $F_{Y_2|\eta_1,p,I}(y_2)$ is the conditional distribution of Y_2 given (η_1, p, I), and so on. (McFadden (1985) and Benkard and Berry (2004) previously considered also using this type of representations for simultaneous equations models.)

By construction, the system

$$\begin{aligned} y_1 &= F^{-1}_{y_1|p,I}(\eta_1) \\ y_2 &= F^{-1}_{y_1|p,I}(\eta_1,\eta_2) \\ &\quad . \quad . \quad . \\ y_{K-1} &= F^{-1}_{y_1|p,I}\left(\eta_1,\eta_2,...,\eta_{K-1}\right) \end{aligned}$$

is observationally equivalent to

$$\begin{aligned} y_1 &= d_1(p,I,\varepsilon_1,...,\varepsilon_{K-1}) \\ y_2 &= d_2(p,I,\varepsilon_1,...,\varepsilon_{K-1}) \\ &\quad . \quad . \quad . \\ y_{K-1} &= d_{K-1}(p,I,\varepsilon_1,...,\varepsilon_{K-1}) \end{aligned}$$

Assuming that (p,I) has an appropriate large support, the connection between both systems can be analyzed using Theorem 3.1 in Matzkin (2005). Denote the mapping from (η,p,I) to y by

$$y = s(\eta,p,I)$$

and the mapping from (y,p,I) to ε by

$$\varepsilon = r(y,p,I)$$

Define the mapping g by

$$\begin{aligned} \varepsilon &= g(\eta,p,I) \\ &= r(s(\eta,p,I),p,I) \end{aligned}$$

Then, by Theorem 3.1 in Matzkin (2005),

$$\left[-\frac{\partial \log\left(f_\eta\left(\eta\right)\right)}{\partial \eta}+\frac{\partial}{\partial \eta}\log\left(\left|\frac{\partial g\left(\eta,p,I\right)}{\partial \eta}\right|\right)\right]\left[\left(\frac{\partial g\left(\eta,p,I\right)}{\partial \eta}\right)^{-1}\frac{\partial g\left(\eta,p,I\right)}{\partial (p,I)}\right]$$

$$= \frac{\partial}{\partial (p,I)}\log\left(\left|\frac{\partial g\left(\eta,p,I\right)}{\partial \eta}\right|\right)$$

Making use of the relationships $g\left(\eta,p,I\right)=r\left(s(\eta,p,I),p,I\right)$ and $y=s(\eta,p,I)$, this expression can be written solely in terms of the unknown elements

$$\frac{\partial r\left(y,p,I\right)}{\partial y},\ \frac{\partial r\left(y,p,I\right)}{\partial x},\ \frac{\partial}{\partial y}\log\left(\left|\frac{\partial r\left(y,p,I\right)}{\partial y}\right|\right),\ \text{and}\ \frac{\partial}{\partial (p,I)}\log\left(\left|\frac{\partial r\left(y,p,I\right)}{\partial y}\right|\right)$$

and the known elements

$$\frac{\partial \log\left(f_\eta\left(\eta\right)\right)}{\partial \eta},\ \frac{\partial s\left(\eta,p,I\right)}{\partial \eta},\ \frac{\partial s\left(\eta,p,I\right)}{\partial (p,I)},\ \frac{\partial}{\partial \eta}\log\left(\left|\frac{\partial s\left(\eta,p,I\right)}{\partial \eta}\right|\right),\ \text{and}\ \frac{\partial}{\partial (p,I)}\log\left(\left|\frac{\partial s\left(\eta,p,I\right)}{\partial \eta}\right|\right)$$

The latter group is known because s is constructed from the conditional distribution functions. One can then use the resulting expression, which contains the elements of both groups, to determine properties of the derivatives of the function r.

2.3. Linear choice sets, no price variation or income variation.

In some cases, we may observe no variation in either prices or income. Only one budget set together with a distribution of choices over it is available for analysis. This is a similar problem to that considered in the hedonic models studied by Ekeland, Heckman, and Nesheim (2004) and Heckman, Matzkin and Nesheim (2002), where observations correspond to only one price function, and consumers and firm locate along that function. Rather than using identification due to variation in prices, they achieve identification exploiting variation in observable characteristics. We next exemplify their method in the consumer demand model. Suppose that $z=(z_1,...,z_{K-1})$ is a vector of observable characteristics. In analogy to the piecewise linear specification in the previous section, suppose that the random utility function is

$$V(y_1,...,y_K,\varepsilon_1,...,\varepsilon_{K-1},z_1,...,z_{K-1}) = U(y_1,...,y_{K-1}) + \sum_{k=1}^{K-1} y_k \cdot (\varepsilon_k + z_k) + y_K$$

Then, the first order conditions for utility maximization are

$$\varepsilon_k = p_k - U_k\left(y_1,...,y_{K-1}\right) - z_k$$

This is identical to the model in Section 2.1.1 where prices vary, after substituting p_k by $p_k - z_k$. Hence, the analysis of identification is analogous to that in Section 2.1. (See also the related literature on equivalence scales (e.g., Lewbel (1989).)

2.4. Nonlinear choice sets

In many situations, budget sets are nonlinear. These can be generated, for example, in the classical consumer demand models by prices that depend on quantities, in the labor supply model by regressive or progressive taxes, or in a characteristics model by feasible combinations of characteristics that can be obtained by existent products. Epstein (1981) provided integrability conditions in this case. Richter (1966), Matzkin (1991), and Yatchew (1985) provided revealed preference results. Hausman (1985) and Bloomquist and Newey (2002), among others, provided methods for estimation of some of these models.

Suppose that the choice problem of a consumer with unobserved ε is to maximize

$$V(y_1,...,y_{K-1},\varepsilon_1,...,\varepsilon_{K-1}) + y_K$$
$$\text{subject to} \quad g(y_1,...,y_K,w) \leq 0$$

where w is an observable vector of variables (e.g, taxes, income) that affects choice sets and is distributed independently of ε, and for all $y_1,...,y_{K-1},w$, g is strictly increasing in y_K. The first order conditions for optimization are, for $k = 1,...,K-1$

$$V_k\left(y_1,...,y_{K-1},\varepsilon_1,...,\varepsilon_{K-1}\right) = s_k\left(y_1,..,y_{K-1},w\right)$$

where $s\left(y_1,..,y_{K-1},w\right) = y_K$ such that $g(y_1,...,y_K,w) = 0$, and where s_k denotes the derivative of s with respect to its $k-th$ coordinate. (We assume that the solution to the FOC is unique and that second order conditions are satisfied.)

This is a multidimensional case of the problem considered by Heckman, Matzkin and Nesheim (2002), when the price function varies with z. Suppose that $DV(y_1,...,y_{K-1},\varepsilon_1,...,\varepsilon_{K-1})$ is invertible in $\varepsilon = (\varepsilon_1,...,\varepsilon_{K-1})$. Let

$$\varepsilon = r\left(y_1,...,y_{K-1},s_k\left(y_1,..,y_{K-1},w\right)\right)$$

Consider an alternative gradient function $D\widetilde{V}(y_1,...,y_{K-1},\varepsilon_1,...,\varepsilon_{K-1})$, which is also invertible in ε. Denote the inverse function of $D\widetilde{V}$ with respect to ε by $\widetilde{r}$, so that

$$\widetilde{\varepsilon} = \widetilde{r}\left(y_1,...,y_{K-1},s_k\left(y_1,..,y_{K-1},w\right)\right)$$

Then, again one can use the results in Matzkin (2005) to determine under what conditions $\widetilde{r}$ is not observationally equivalent to r. Constructive identification can be obtained by specifying the utility function as

$$U(y_1,...,y_{K-1}) + \sum_{k=1}^{K-1} \varepsilon_k y_k + y_K$$

In this case, the first order conditions become

$$\varepsilon = s_k\left(y_1,..,y_{K-1},w\right) - U_k\left(y_1,...,y_{K-1}\right)$$

Specifying the value of the gradient $DU = (U_1,...,U_{K-1})$ of U at a point $\overline{y}$ and assuming that for each y and s_k, the value of $s_k\left(y_1,..,y_{K-1},w\right)$ has an appropriate large support, as w varies, one can follow a method similar to that discussed in Sections 1.1 and 1.2 to directly recover f_ε and $DU(y) \equiv (U_1(y),...,U_{K-1}(y))$ from the distribution of the observable variables.

If w were not observed varying, this would become a situation like the single market observations in Ekeland, Heckman and Nesheim (2004) and Heckman, Matzkin, and Nesheim (2002). Rather than exploiting exogenous variation in prices, one exploits exogenous variation in the observable characteristics of the consumers. Suppose, for example, that the choice problem of the consumer with unobserved ε is to maximize

$$U(y_1,...,y_{K-1}) + y_K + \textstyle\sum_{k=1}^{K-1}(z_k+\varepsilon_k)y_k$$
$$\text{subject to} \quad g(y_1,...,y_K) \le 0$$

The first order conditions for optimization are, for $k = 1, ..., K-1$,

$$U_k\left(y_1, ..., y_{K-1}\right) + z_k + \varepsilon_k = s_k\left(y_1, .., y_{K-1}\right)$$

where $s\left(y_1, .., y_{K-1}, z\right) = y_K$ such that $g(y_1, ..., y_K, z) = 0$. Hence, variation in the observable characteristics z can be used to recover the gradient of U and the density of ε.

2.5. Restrictions on the distribution of choices

Given a distribution of choices, over any type of choice sets, one may ask what restrictions optimization places on this distribution. This question was studied by McFadden and Richter (1970, 1990), Falmagne (1978), Cohen (1980), and Barbera and Pattanaik (1986), among others, (see Fishburn (1998)). Most recently, the issue has been studied by McFadden (2005). In particular, consider a discrete set of alternatives. Let $\widetilde{R}$ denote the set of all possible orderings of those alternatives. The following is defined in McFadden and Richter (1991):

Axiom of Stochastic Revealed Preference: For every finite list $\{(B_1, C_1), ..., (B_k, C_k)\}$ of subsets of alternatives, $C_1, ..., C_k$, and choices from those alternatives, $B_1, ..., B_k$,

$$\sum_{i=1}^{k} \pi_{B_i}(C_i) \leq Max\left\{\sum_{i=1}^{k} \alpha(R, B_i, C_i) | R \in \widetilde{R}\right\}$$

where $\pi_{B_i}(C_i)$ is the probability of choosing B_i from choice set C_i; $\alpha(R, B_i, C_i) = 1$ if ordering R selects B_i from C_i and $\alpha(R, B_i, C_i) = 0$ otherwise.

McFadden and Richter (1970) showed that this axiom is equivalent to the existence of a probability distribution over the space of orderings $\widetilde{R}$. They proved the following:

Theorem (McFadden and Richter (1970))**:** *The Axiom of Stochastic Revealed Preference is equivalent to the existence of a probability ξ on the space of ordering such that for all orderings $R \in \widetilde{R}$, all choice sets C, and all subsets B of C*

$$\pi_B(C) = \xi\left(\left\{R \in \widetilde{R} | R \text{ selects an element of } B \text{ from } C\right\}\right)$$

McFadden (2005) has generalized this result to non-discrete budget sets. To provide an example of what this result means, consider two budget sets in R_+^2 that intersect. Restrict the set of random preferences to be such that each random preference admits a strictly monotone, strictly concave, and twice differentiable utility function, and the relationship between ε and the choice generated by ε is 1-1. Let A denote the set of points in the first budget line, C_1, that are included in the second budget set. Let B denote the set of points in the second budget line, C_2, that are included in the first budget set. The axiom implies that

$$\pi_A(C_1) + \pi_B(C_2) \leq 1.$$

That is, the probability of the set of orderings that choose an element of C_1 plus the probability of the set of orderings that choose an element of C_2 has to be not larger than 1. If this sum were larger than 1, it would imply that the probability of the set of ordering that choose an element of C_1 in A and an element of C_2 in B is strictly positive.

3. Discrete choice

In many situations in economics, consumer's feasible set is only a finite sets of alternatives. Discrete Choice Models (McFadden (1974)) provide the appropriate framework to analyze such situations. In these models, each alternative is characterized by a vector of characteristics. The consumer's utility of each alternative is given by a function, which may depend on observable characteristics of the consumer and the alternative, and on unobservable characteristics of either the consumer, the alternative, or both. It is assumed that the alternative chosen by the consumer is the one that provides the highest level of utility. In a standard textbook example, the available alternatives are to commute to work by either car or bus. The observable characteristics of each means of transportation are its time and its cost. Comfort might be one of the unobservable characteristics. The observable characteristic of the consumer are the number of cars owned by the commuter and his income.

Let J denote the set of feasible alternatives. Let $V_j(s, z_j, x_j, \omega)$ denote the utility for alternative j, where s denotes a vector of observable characteristics of the consumer, x_j denotes a vector of observable attributes of alternative j, z_j denotes an observable variable which denotes another attribute of the alternative, and ω is an unobservable random vector. Let $y = (y_1, ..., y_J)$ be defined by

$$y_j = \left\{ \begin{array}{lll} 1 \quad \text{if} & V_j(s, z_j, x_j, \omega) > V_k(s, z_k, x_k, \omega) & \text{for all } k \neq j \\ 0 & & \text{otherwise} \end{array} \right\}$$

Then, the conditional choice probabilities, for each $j = 1, ..., J$ is

$$\Pr(\{y_j = 1 \mid s, x, z\}) = \Pr(\{\omega \mid V_j(s, z_j, x_j, \omega) > V_k(s, z_k, x_k, \omega) \quad \text{for all } k \neq j \ \})$$

Since the choice probabilities of each alternative depend only on the differences between the utilities of the alternatives, only those differences can be identified. Hence, for simplicity, we will normalize $V_J(s, z_J, x_J, \omega)$ equal to 0 for all (s, z_J, x_J, ω). Then,

$$\Pr(\{y_J = 1 \mid s, x, z\}) = \Pr(\{\omega \mid 0 > V_k(s, z_k, x_k, \omega) \quad \text{for all } k \neq J \ \})$$

(We assume that the probabiloty of ties is zero.)

3.1. Subutilities additive in the unobservables

Assume that $\omega = (\omega_1, ..., \omega_{J-1})$, each V_j depends only on one coordinate, ω_j of ω, and that each ω_j is additive. Then, under certain restrictions, one can identify the V_j functions and the distribution of ω nonparametrically. A result of this type, considered in Matzkin (1992, 1993,1994), is where, for each j, V_j is specified as

$$(3.1) \quad V_j(s, x_j, z_j, \omega) = z_j + v_j(s, x_j) + \omega_j,$$

where v_j is a nonparametric function. Matzkin (1992, 1993, 1994) requires that $(z_1, ..., z_{J-1})$ has an everywhere positive density, conditional on $(s, x_1, ..., x_{J-1})$ and that $(\omega_1, ..., \omega_{J-1})$ is distributed independently of $(s, z_1, ..., z_{J-1}, x_1, .., x_{J-1})$. In addition, the vector of functions $(v_1, ..., v_{J-1})$ is normalized by requiring that at a point, $(\overline{s}, \overline{x}_1, ..., \overline{x}_{J-1})$, and for some values $(\alpha_1, ..., \alpha_{J-1})$,

$$(3.2) \quad v_j(\overline{s}, \overline{x}_j) = \alpha_j$$

Matzkin (1992, Example 3) shows that the functions v_j and the distribution of $(\omega_1, ..., \omega_{J-1})$ are identified nonparametrically. Lewbel (2000) shows that to identify the distribution of $(v_1(s, x_1) + \omega_1, ..., v_J(s, x_J) + \omega_J)$ given $(s, x_1, ..., x_J)$, it suffices to require that $(\omega_1, ..., \omega_{J-1})$ is distributed independently of $(z_1, ..., z_{J-1})$, conditional on $(s, x_1, .., x_{J-1})$. To separately identify $(v_1(s, x_1), ..., v_{J-1}(s, x_{J-1}))$ from the distribution of $(\omega_1, ..., \omega_J)$, Lewbel imposes linearity of the functions v and a zero covariance between ω and $(s, x_1, ..., x_J)$. Alternatively, Lewbel uses instruments for $(s, x_1, ..., x_J)$.

In the past, the identification and estimation of these models, was analyzed using either parametric or semiparametric methods. McFadden (1974, 1981) specified each function V_j as linear in (s, z_j, x_j) and additive in ω_j. The distribution of ω was assumed to be known up to a finite dimensional parameter. Manski (1975) developed an estimation method for the parameters of the linear function that did not require a parametric specification for the distribution of ω. Later, Cosslett (1983), Manski (1985), Powell Stock and Stoker (1989), Horowitz (1992), Ichimura (1993), and Klein and Spady (1993), among others, developed other methods that did not require a parametric specification for the distribution of ω. Matzkin (1991) developed methods in which the distribution of $(\omega_1, ..., \omega_{J-1})$ is parametric and the functions V_j are additive in ω_j but otherwise nonparametric. (See Briesch, Chintagunta, and Matzkin (2002) for an application of this latter method.)

To see how identification is achieved, consider the specification in (3.1) with the normalization in (3.2) and the assumption that $\omega = (\omega_1, ..., \omega_{J-1})$ is distributed independently of $(s, z, x) = (s, z_1, ..., z_J, x_1, ..., x_{J-1})$. Then,

$$\begin{aligned} & \Pr(\{y_J = 1 | \ s, z_1, ..., z_{J-1}, x_1, ..., x_{J-1}\}) \\ = \ & \Pr(\{ \ 0 > z_j + v_j(s, x_j) + \omega_j \quad \text{for all } j \neq J \mid s, z_1, ..., z_{J-1}, x_1, ..., x_{J-1}\}) \\ = \ & \Pr(\{ \ \omega_j < -z_j - v_j(s, x_j) \quad \text{for all } j \neq J \ \mid s, z_1, ..., z_{J-1}, x_1, ..., x_{J-1} \ \}) \end{aligned}$$

Since independence between ω and (s, z, x) implies that ω is independent of z, conditional on (s, x), it follows from the last expression that fixing the values of (s, x) and varying the values of z, we can identify the distribution of ω. Specifically, by letting $(s, x) = (\overline{s}, \overline{x}_1, ..., \overline{x}_{J-1})$, it follows that

$$\begin{aligned}
& \Pr\left(\{y_J = 1 | \ \overline{s}, z_1, ..., z_{J-1}, \overline{x}_1, ..., \overline{x}_{J-1}\}\right) \\
= \ & \Pr\left(\{ \ \omega_j < -z_j - \alpha_j \quad \text{for all } j \neq J \ \ | \ \overline{s}, z_1, ..., z_{J-1}, \overline{x}_1, ..., \overline{x}_{J-1}\}\right)
\end{aligned}$$

This shows that the distribution of $(\omega_1, ..., \omega_{J-1})$, conditional on $(s, x) = (\overline{s}, \overline{x}_1, ..., \overline{x}_{J-1})$ is identified. Since ω is distributed independently of (s, x), the marginal distribution of ω is the same as the conditional distribution of ω. Hence, for any $(t_1, ..., t_{J-1})$,

$$F_{\omega_1,...,\omega_{J-1}}(t_1, ..., t_{J-1}) = \Pr\left(\{y_J = 1 | \ \overline{s}, z_1 = -\alpha_1 - t_1, ..., z_{J-1} = -\alpha_{J-1} - t_{J-1}, \overline{x}_1, ..., \overline{x}_{J-1}\}\right)$$

Once the distribution of the additive random terms is identified, one can identify the nonparametric functions V_j using, for example, the identification results in Matzkin (1991), which assume that the distribution of ω is parametric.

3.2. Subutilities nonadditive in the unobservables

Matzkin (2005) considered a model where the vector of unobservables, $\omega = (\omega_1, ..., \omega_{J-1})$, entered in possible nonadditive ways in each of the V_j functions. The specification was, for each j,

$$(3.3) \quad V_j(s, x_j, z_j, \omega) = z_j + v_j(s, x_j, \omega),$$

where v_j is a nonparametric function, and ω is assumed to be distributed independently of $(s, z_1, ..., z_{J-1}, x_1, .., x_{J-1})$. The analysis of identification makes use of the arguments in Matzkin (1992, 1993, 1994), Briesch, Chintagunta and Matzkin (1997), and Lewbel (1998, 2000), together with the results about identification in nonparametric simultaneous equations models. Define U_j for each j by

$$U_j = v_j(s, x_j, \omega).$$

Since ω is distributed independently of (s, z, x), $(U_1, ..., U_{J-1})$ is distributed independently of z given (s, x). Hence, using the arguments in Lewbel (1998, 2000), from the probability of choosing alternative J, or any other alternative, conditional on (s, x), one can recover the distribution of $(U_1, .., U_{J-1})$ given (s, x). Since the distribution of the observable variables (s, x) is observable, this is equivalent to observing the joint distribution of the dependent variables $(U_1, ..., U_{J-1})$ and the explanatory variables, $(s, x_1, ..., x_J)$, in the system of equations

$$\begin{aligned} U_1 &= v_1 (s, x_1, \omega_1, ..., \omega_{J-1}) \\ U_2 &= v_2 (s, x_2, \omega_1, ..., \omega_{J-1}) \\ &\quad . \quad . \quad . \\ U_{J-1} &= v_{J-1} (s, x_{J-1}, \omega_1, ..., \omega_{J-1}) \end{aligned}$$

The identification and estimation of the functions $(v_1, ..., v_{J-1})$ in this system can be analyzed using the methods in Matzkin (2005).

For a simple example, that generalizes the model in 3.1 to functions that are nonadditive in one unobservable, suppose that each function V_j is of the form

$$\text{(3.4)} \quad V_j (s, x_j, z_j, \omega) = z_j + v_j (s, x_j, \omega_j),$$

where each nonparametric functions v_j is strictly increasing in ω_j. Then, assuming the ω is distributed independently of (s, z, x), one can identify the joint distribution of $(U_1, ..., U_{J-1}; s, x_1, ..., x_{J-1})$, where

$$\begin{aligned} U_1 &= v_1 (s, x_1, \omega_1) \\ U_2 &= v_2 (s, x_2, \omega_2) \\ &\quad . \quad . \quad . \\ U_{J-1} &= v_{J-1} (s, x_{J-1}, \omega_{J-1}) \end{aligned}$$

Each of these v_j functions can be identified from the distribution of (U_j, s, x_j) after either a normalization or a restriction. For example, one could require that each function v_j is separable into a known function of one coordinate of (s, x_j) and ω_j, and the value of v_j is known at one point, or, one could specify a marginal distribution for ω_j.

One could also consider situations where the number of unobservables is larger than $J - 1$. Briesch, Chintagunta, and Matzkin (1997), for example, consider a situation where the utility function depends on an unobserved heterogeneity variable, θ. They specify, for each j

$$V(j, s, x_j, z_j, \theta, \omega_j) = z_j + v(j, s, x_j, \theta) + \omega_j$$

where θ is considered an unobservable heterogeneity parameter.

Ichimura and Thompson (1998) consider the model

$$V(j, s, x_j, z_j, \theta) = z_j + \theta\ x_j$$

where θ is an unobservable vector, distributed independently of (z, x), with an unknown distribution. Using Matzkin (2003, Appendix A), one can modify the model in Ichimura and Thompson (1998) to allow the random coefficients to enter in a nonparametric way. This can be done, for example, by specifying, for each j, that $V(j, s, x_j, z_j, \theta_j, \omega_j)$ is of the form

$$V(j, s, x_j, z_j, \theta_j, \omega_j) = z_j + \sum_{k=1}^{K} m_k\left(x_{j(k)}, \theta_{j(k)}\right) + \omega_j$$

where the nonparametric functions m_k are a.e. strictly increasing in θ_j, the θ_j are independently distributed among them and from (z, x), and for some $\overline{x}_j, \widetilde{x}_j$,

$$\begin{aligned} m_k\left(\widetilde{x}_{j(k)}, \theta_j\right) &= 0 \text{ for all } \theta_j, \text{ and} \\ m_k\left(\overline{x}_{j(k)}, \theta_j\right) &= \theta_j \end{aligned}$$

In the random coefficient model of Ichimura and Thompson (1998), these conditions are satisfied at $\widetilde{x}_{j(k)} = 0$ and $\overline{x}_{j(k)} = 1$.

4. Choices as explanatory variables

In many models, choices appear as explanatory variables. When this choice depends on unobservables that affect the dependent variable of interest, we need to deal with an endogeneity problem, to be able to disentangle the effect of the observable choice variable from the effect of the unobservable variables. Suppose, for example, that the model of interest is

$$Y = m\,(X, \varepsilon)$$

where Y denotes output of a worker that possesses unobserved ability ε and X is the amount of hours of work. If the value of X is determined independently of ε, then one could analyze the identification of the function m and of the distribution of ε using Matzkin (1999), in a way similar to that used in the analysis of demand in Section 2.1. In particular, assuming that m is strictly increasing in ε, it follows by the independence between X and ε that for all x, e

$$F_{\varepsilon}(e) = F_{Y|X=x}\,(m(x,e))$$

Suppose, in addition, that $f_{\varepsilon}(e) > 0$. Then, at $X = x$, the conditional distribution of Y given x when $Y = m(x, e)$ is strictly increasing, locally. It follows that

$$m(x,e) = F^{-1}_{Y|X=x}\,(F_{\varepsilon}(e))$$

and that a change in the value of X from x to x' causes a change in the value of Y from $y_0 = m(x, e)$ to $y_1 = m(x, e)$ equal to:

$$y_1 - y_0 = F^{-1}_{Y|X=x'}\left(F_{Y|x=x}(y_0)\right) - y_0$$

Suppose, however, that the amount of hours of work is chosen by the worker, as a function of his ability. In particular, suppose that, for some ξ,

$$X = v(\varepsilon, \xi)$$

where v is strictly increasing in ε. Then, there exists an inverse function r such that

$$\varepsilon = r\,(X, \xi)$$

The change in output, for a worker of ability ε, when the amount of X changes exogenously from x to x' is

$$y_1 - y_0 = m(x', r(x,\xi)) - m(x, r(x,\xi))$$

Altonji and Matzkin (1997, 2005), Altonji and Ichimura (2000), Imbens and Newey (2001, 2003), Chesher (2002, 2003), and Matzkin (2003, 2004) consider local and global identification of m and of average derivatives of m, in nonseparable models, using conditional independence methods. Altonji and Matzkin (1997) also propose a method that uses shape restrictions on a conditional distribution of ε. All of these methods are based upon using additional data. Most are based on a control function approach, following the parametric methods in Heckman (1976, 1978, 1980), and later works by Heckman and Robb (1985), Blundell and Smith (1986, 1989), and Rivers and Vuong (1988). Control function approaches for nonparametric models that are additive on an unobservable have been studied by Newey, Powell and Vella (1999), Ng and Pinske (1995), and Pinske (2000).

The shape restricted method in Altonji and Matzkin (1997, 2005) assumes that there exists an external variable, Z, such that for each value x of X there exists a value z of Z such that for all e

$$f_{\varepsilon|X=x,Z=z}(e) = f_{\varepsilon|X=z,Z=x}(e)$$

In the above example, Z could be the amount of hours of that same worker at some other day. The assumption is that the distribution of ability, ε, given that the hours of work are x today and z at that other period is the same distribution as when the worker's hours today are z and the worker's hours at the other period are x. Assuming that m is strictly increasing in ε, this exchangeability restriction implies that

$$F_{Y|X=x,Z=z}(m(x,e)) = F_{Y|X=z,Z=x}(m(z,e))$$

This is an equation in two unknowns, $m(x,e)$ and $m(z,e)$. Altonji and Matzkin (1997) use the normalization that at some $\overline{x}$, and for all e,

$$m\left(\overline{x}, e\right) = e$$

This then allows one to identify $m(x,e)$ as

$$m(x,e) = F^{-1}_{Y|X=x,Z=\overline{x}}\left(F_{Y|X=\overline{x},Z=x}(e)\right)$$

The distribution $F_{\varepsilon|X=x}$ can next be identified by

$$F_{\varepsilon|X=x}(e) = F_{Y|X=x}(F^{-1}_{Y|X=x,Z=\overline{x}}\left(F_{Y|X=\overline{x},Z=x}(e)\right))$$

Altonji and Matzkin (1997, 2005) also propose a method to estimate average derivatives of m using a control variable Z satisfying the conditional independence assumption that

$$(4.1) \qquad F_{\varepsilon|X,Z}(e) = F_{\varepsilon|Z}(e)$$

Their object of interest is the average derivative of m with respect to x, when the distribution of ε given X remains unchanged:

$$\beta(x) = \int \frac{\partial m(x,e)}{\partial x}\ f_{\varepsilon|X=x}(e)\ de$$

They show that under their conditional independence assumption, $\beta(x)$ can be recovered as a functional of the distribution of the observable variables (Y,X,Z). In particular,

$$\beta(x) = \int \frac{\partial E\left(Y|X=x,Z=z\right)}{\partial x}\ f(z|x)\ dz$$

where $E\left(Y|X=x,Z=z\right)$ denotes the conditional expectation of Y given $X=x, Z=z$, and $f(z|x)$ denotes the density of Z given $X=x$. Matzkin (2003) showed how condition (4.1) can also be used to identify m and the distribution of ε given X, using the normalization that at some $\overline{x}$, and for all ε

$$m\left(\overline{x},\varepsilon\right) = \varepsilon$$

Blundell and Powell (2000, 2003) and Imbens and Newey (2001, 2003) have considered situations where the control variable Z, satisfying (4.1) is estimated rather than observed. Blundell and Powell considered the Average Structural Function

$$\int m(x, \varepsilon)\ f_{\varepsilon}(\varepsilon)\ d\varepsilon$$

as well as functionals of it and average derivatives. Imbens and Newey considered the Quantile Structural Function, defined as

$$m\left(x, q_{\varepsilon}(\tau)\right)$$

where $q_{\varepsilon}(\tau)$ is the $\tau - th$ quantile of ε, as well as the Average Structural Function, functionals of both functions, and average derivatives. (See Imbens (2006).) Condition (4.1) together with strict monotonicty of m in ε implies that

$$F_{\varepsilon|Z}(\varepsilon) = F_{Y|X,Z}(m(x, \varepsilon))$$

Multiplying both sides of the equation by $f(z)$ and integrating with respect to z, one gets

$$F_{\varepsilon}(\varepsilon) = \int F_{Y|X,Z}(m(x, \varepsilon))\ f(z)\ dz$$

which can be solved for $m(x, \varepsilon)$ once one specifies the marginal distribution of ε.

The above approaches are based on conditional independence, given a variable, Z, which can be either observed or estimated. One may wonder, however, how to determine whether any given variable, Z, can be used to obtain conditional independence. Imbens and Newey (2001) provides one such example. They consider the model

$$\begin{aligned} Y &= m(X, \varepsilon) \\ X &= s(Z, \eta) \end{aligned}$$

where s is strictly increasing in η, m is strictly increasing in ε, and Z is jointly independent of (ε, η). The latter assumption implies that ε is independent of X conditional on η, because conditional conditional on η, X is a function of Z, which is independent of ε. Since the function s can be identified (Matzkin (1999)), one can estimate η, and use it to identify m. Chesher (2002, 2003) considered a model similar to the one in Imbens and Newey, and showed that to identify local derivatives of m, a local quantile insensitivity condition was sufficient. (See also Ma and Koenker (2004).)

Matzkin (2004) shows that reversing the roles in the second equation in Imbens and Newey can, in some situations, considerably lessen the requirements for global identification. The model is

$$\begin{aligned} Y &= m(X, \varepsilon) \\ X &= s(Z, \eta) \end{aligned}$$

where for one value, $\overline{z}$ of Z, ε is independent of η, conditional on $Z = \overline{z}$. It is assumed that m is strictly increasing in ε, $F_{\varepsilon|Z=\overline{z}}$ and $F_{X|Z=\overline{z}}$ are strictly increasing, and for each value x of X, $F_{\varepsilon|X=x,Z=\overline{z}}$ is strictly increasing. Under these assumptions, Matzkin (2004) showed that if η is independent of ε conditional on $Z = \overline{z}$, then for all x, e

$$m(x, e) = F^{-1}_{Y|X=x,Z=\overline{z}}\left(F_{\varepsilon|Z=\overline{z}}(e)\right) \quad \text{and}$$

$$F_{\varepsilon|X=x}(e) = F_{Y|X=x}\left(F^{-1}_{Y|X=x,Z=\overline{z}}\left(F_{\varepsilon|Z=\overline{z}}(e)\right)\right)$$

This result establishes the global identification of the function m and the distribution of (X, ε), up to a normalization on the conditional distribution $F_{\varepsilon|Z=\overline{z}}$. If, for example, we normalized the distribution of ε conditional on $Z = \overline{z}$ to be $U(0,1)$, then, for all x and all $e \in (0,1)$

$$m(x, e) = F^{-1}_{Y|X=x,Z=\overline{z}}(e) \qquad \text{and} \qquad F_{\varepsilon|X=x}(e) = F_{Y|X=x}\left(F^{-1}_{Y|X=x,Z=\overline{z}}(e)\right)$$

Endogeneity in discrete choice models can be handled similarly. In particular, Altonji and Matzkin (1997, 2005)'s average derivative method can be applied to the case where Y is discrete as well as when Y is continuous. Blundell and Powell (2003) derive an estimator for a semiparametric binary response model with endogenous regressors, using a control function approach. Similarly to Blundell and Powell (2003), but with less structure, one can consider a nonparametric binary response model, as in Matzkin (2004). Consider the model

$$Y_1 = \left\{ \begin{array}{lcc} 1 & if & X_0 + v\left(X_1, Y_2, \xi\right) \geq \varepsilon \\ 0 & & otherwise \end{array} \right\}$$

where Y_1, Y_2, and X are observable and ξ and ε are unobservable. Suppose that ξ is not distributed independently of Y_2. Let

$$W = v\left(X_1, Y_2, \xi\right)$$

and assume that for some unknown function s, observable $\widetilde{Y}_2$, and unobservable η

$$Y_2 = s\left(\widetilde{Y}_2, \eta\right)$$

where η is independent of ξ conditional on $\widetilde{Y}_2$. Following arguments similar to those described in Section 3, one can identify the distribution of W given $\left(X_1, Y_2, \widetilde{Y}_2\right)$, from the distribution of Y_1 given $\left(X_0, X_1, Y_2, \widetilde{Y}_2\right)$, using X_0 as a special regressor (Lewbel (2000)). The system

$$W = v\left(X_1, Y_2, \xi\right)$$

$$Y_2 = s\left(\widetilde{Y}_2, \eta\right)$$

is a nonparametric triangular system whose identification can be analyzed using Chesher (2003), Imbens and Newey (2001, 2003), Matzkin (2004), or using methods for nonparametric simultaneous equations models, such as the ones mentioned in Section 3.

5. Conclusions

This paper has surveyed some of the current literature on nonparametric identification of models with unobserved heterogeneity. We considered models where choices are the dependent variables, such as the classical consumer demand model and discrete choice models, and models where choices are explanatory variables, such as the model where years of education is an explanatory variable in the determination of wages.

We concentrated on models where the unobservable random terms, representing unobserved heterogeneity, entered in the relationships in nonadditive ways. In some of the models, these unobservable could be considered as independently distributed from the observable explanatory variables. In others, they were conditionally independent of the explanatory variables. In some, such as in demand models generated from multivariate unobservables, we dealt with nonparametric models of simultaneous equations with nonadditive unobservables.

Several variations in the support of the data were considered. For the classical consumer demand model, in particular, we presented existent results that can be used when observed prices and income are continuously distributed, when only observed income is continuously distributed, and when neither prices nor income vary over observations. Linear and nonlinear budget constraints were considered.

6. References

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

AI, C. and X. CHEN (2003), "Efficient Estimation of Models with Conditional Moments Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795-1843.

ALTONJI, J.G. AND H. ICHIMURA (2000): "Estimating Derivatives in Nonseparable Models with Limited Dependent Variables," mimeo.

ALTONJI, J.G. AND R.L. MATZKIN (1997): "Panel Data Estimators for Nonseparable Models with Endogenous Regressors," mimeo, Northwestern University, appeared as NBER Working paper T0267 (March 2001).

ALTONJI, J.G. AND R.L. MATZKIN (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica,* Vol. 73, No. 3, pp. 1053-1102.

ATHEY, S and G. IMBENS (2006): "Identification and Inference in Nonlinear Difference-in-Difference Models," *Econometrica,* 74 (2), 431-497.

BARTEN, A.P. (1968), "Estimating Demand Equations," *Econometrica,* Vol. 36, No. 2, 213-251.

BARBERA, S. and P.K. PATTANAIK (1986) "Flamange and the Rationalizability of Stochastic Choices in Terms of Random Orderings," Econometrica, Vol. 54, No. 3, pp. 707-715.

BECKERT, W. and R. BLUNDELL (2005), "Heterogeneity and the Nonparametric Analysis of Consumer Choice: Conditions for Invertibility," CEMMAP Working Paper, CWP 09/05.

BENKARD, C.L. and S. BERRY (2004, 2006) "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on B. Brown (1983) and Roehrig (1988)," Cowles Foundation Discussion Paper #1482 (2004), published in *Econometrica,* Vol. 74, No. 5, pp. 1429-1440 (2006).

BLOOMQUIST, S. and W. NEWEY (2002) "Nonparametric Estimation with Nonlinear Budget Sets," *Econometrica*, 70(6), 2455-2480.

BLUNDELL, R., M. BROWNING and I. CRAWFORD (2003), "Nonparametric Engel Curves and Revealed Preference" *Econometrica,* Vol.71, No.1. January, 205-240.

BLUNDELL, R., M. BROWNING and I. CRAWFORD (2004), "Best Nonparametric Bounds on Demand Responses," Walras-Bowley :ecture, IFS Working Paper, 05/03.

BLUNDELL, R., M. BROWNING and I. CRAWFORD (2005), "Improving Revealed Preference Bounds on Demand Responses," mimeo, IFS.

BLUNDELL, R., X. CHEN, and D. KRISTENSEN (2003), "Semiparametric Engel Curves with Endogenous Expenditure," CEMMAP Working Paper, 15/03.

BLUNDELL, R., D. KRISTENSEN, and R.L. MATZKIN (2005) "Stochastic Demand and Revealed Preference", mimeo.

BLUNDELL, R. and R.L. MATZKIN (2005) "Nonseparable Demand", mimeo.

BLUNDELL, R. and J. L. POWELL (2000) "Endogeneity in Nonparametric and Semiparametric Regression Models," Invited Lecture at the Eigth World Congress of the Econometric Society, Seattle, Washington, USA.

BLUNDELL, R. and J. L. POWELL (2003) "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics, Theory and Applications, Eighth World Congress,* Volume II, edited by M. Dewatripont, L.P. Hansen, and S.J.Turnovsky, Cambridge University Press, Cambridge, U.K.

BLUNDELL, R. and J. L. POWELL (2004) "Endogeneity in Semiparametric Binary Response Models," *Review of Economics Studies,* Volume 71, No. 3, pp. 655-679.

BLUNDELL, R. and R. SMITH (1986), "An Exogeneity Test for a Simultaneous Equation Tobit Model with and Application to Labor Supply," *Econometrica*, 54, 3, pp. 679-686.

BLUNDELL, R. and R. SMITH (1989), "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models," *Review of Economic Studies,* 56, 37-58.

BRIESCH, R., P. CHINTAGUNTA, and R. MATZKIN (1997) "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," mimeo, Northwestern University.

BRIESCH, R., P. CHINTAGUNTA, and R. MATZKIN (2002) "Semiparametric Estimation of Choice Brand Behavior," *Journal of the American Statistical Association,* Vol. 97, No. 460, Applications and Case Studies, pp. 973-982.

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

BROWN, D.J. and C. CALSAMIGLIA (2003) "The Strong Law of Demand," CFDP #1399, Yale University.

BROWN, D.J. and R.L. MATZKIN (1998) "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand," Cowles Foundation Discussion Paper #1175.

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

BROWN, B.W. and M.B. WALKER (1995) "Stochastic Specification in Random Production Models of Cost-Minimizing Firms", *Journal of Econometrics,* 66, 1, p. 175-205..

CHERNOZHUKOV, V. AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.

CHERNOZHUKOV, V., G. IMBENS, AND W. NEWEY (2004): "Instrumental Variable Identification and Estimation of Nonseparable Models via Qualtile Conditions," mimeo, Department of Economics, M.I.T., forthcoming in *Journal of Econometrics.*

CHESHER, A. (2002) "Local Identification in Nonseparable Models," CeMMAP Working Paper CWP 05/02.

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, Vol. 71, No. 5, pp. 1405-1441.

CHESHER, A. (2006): "Identification in Non-additive Structural Functions," presented at the Invited Symposium on Nonparametric Structural Models, 9th World Congress of the Econometric Society, London 2005.

COHEN, M. (1980) "Random Utility Systems - the Infinite Case," *Journal of Mathematical Psychology,* 18, 52-72.

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

DAROLLES, S., J.P. FLORENS, and E. RENAULT (2002), "Nonparametric Instrumental Regression," mimeo, IDEI, Toulouse.

EKELAND, I., J.J. HECKMAN and L. NESHEIM (2004) "Identification and Estimation of Hedonic Modles," *Journal of Political Economy,* 112 (S1), S60-S109.

EPSTEIN, L.G. (1981) "Generalized Duality and Integrability," *Econometrica,* Vol. 49, No. 3, pp. 655-678.

FALMANGE, J. (1978) "A Representation Theorem for Finite Random Scale Systems," *Journal of Mathematical Psychology,* 18, 52-72.

FISHBURN, P. (1998) "Stochastic Utility," in *Handbook of Utility Theory,* by S. Barbera, P. Hammond, and C. Seidl (eds.), Kluwer, 273-320.

HALL, P. and J.L. HOROWITZ (2003), "Nonparametric Methods for Inference in the Presence of Instrumental Variables," mimeo, Northwestern University.

HAUSMAN, J. (1985) "The Econometrics of Nonlinear Choice Sets," Econometrica, Vol. 53, No. 6, pp. 1255-1282.

HECKMAN, J.J. (1974) "Effects of Day-Care Programs on Women's Work Effort," *Journal of Political Economy,* 82, pp. 136-163.

HECKMAN, J. (1976); "Simultaneous Equations Models with Continuous and Discrete Endogenous Variables and Structural Shifts," in *Studies in Nonlinear Estimation*, edited by S. Goldfeld and R. Quandt; Cambridge, Mass.: Ballinger.

HECKMAN, J (1978): "Dummy Endogenous Variables in a Simultaneous Equations System," *Econometrica*, 46, 931-61.

HECKMAN, J. (1980); "Addendum to Sample Selection Bias as a Specification Error," in *Evaluation Studies*, vol. 5, edited by E. Stromsdorfer and G. Farkas. San Francisco: Sage.

HECKMAN, J.J., R.L. MATZKIN and L. NESHEIM (2002) "Nonparametric Estimation of Nonadditive Hedonic Models," mimeo, UCL.

HECKMAN, J., and R. ROBB (1985), "Alternative Methods for Evaluating the Impacts of Interventions," in J.J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data,* Economeric Society Monograph 10, Cambridge: Cambridge University Press.

HECKMAN, J. and B. SINGER (1984) "A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica,* 52, 271-320.

HECKMAN, J.J. and R. WILLIS (1977) "A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women," *Journal of Political Economy*, Vol. 87, No. 1, pp. 197-201.

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

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

ICHIMURA, H. (1993) "Semiparametric Least Squares (SLS) ad Weighted SLS Estimation of Single Index Models," *Journal of Econometrics,* 58, 71-120.

ICHIMURA, H. and T.S. THOMPSON (1998) "Maximum Likelihood Estimatin of a Binary Choice Model with Random Coefficients of Unknown Distribution," *Journal of Econometrics,* Vol. 86, No. 2, pp 269-295.

IMBENS, G.W. (2006): "Nonadditive Models with Endogenous Regressors," presented at the Invited Symposium on Nonparametric Structural Models, 9th World Congress of the Econometric Society, London 2005.

IMBENS, G.W. AND W.K. NEWEY (2001, 2003) "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," mimeo, UCLA.

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

KOENKER, R. (2005) *Quantile Regression,* Econometric Society Monograph Series, Cambridge University *Press.*

LANCASTER, T. (1979) "Econometric Methods for the Analysis of Unemployment" *Econometrica,* 47, 939-956.

LEWBEL, A. (1989) "Identification and Estimation of equivalence Scales under Weak Separability," *The Review of Economic Studies,* Vol. 56, No. 2, 311-316.

LEWBEL, A. (1998) "Semiparametric Latent Variables Model Estimation with Endogenous or Missmeasured Regressors," *Econometrica,* Vol. 66, No. 1, pp. 105-121.

LEWBEL, A. (2000) "Semiparametric Qualitative Response Model Estimation with Unknown Hetersokedasticity and Instrumental Variables," *Journal of Econometrics*, 97, 145-177.

LEWBEL, A. (2001) "Demand Systems With and Without Errors," *American Economic Review,* 91, pp. 611-618.

MA, L. and R.W. KOENKER (2004) "Quantile Regression Methods for Recursive Structural Equation Models," CeMMAP Working Paper WP 01/04, forthcoming in *Journal of Econometrics.*

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

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

MANSKI, C.F. (1985) "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics,* Vol. 27, No. 3, pp. 313-333.

MAS-COLELL, A. (1977) "The Recoverability of Consumers' Preferences from Market Demand Behavior," *Econometrica,* Vol. 45, No. 6, pp. 1409-1430.

MAS-COLELL, A. (1978) "On Revealed Preference Analysis," *Review of Economic Studies,* Vol. 45, No. 1, pp. 121-131.

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

MATZKIN, R.L. (1991b) "Axioms of Revealed Preference for Nonlinear Choice Sets," *Econometrica,* Vol. 59, No. 6, pp. 1779-1786.

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

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

MATZKIN, R.L. (1994) "Restrictions of Economic Theory in Nonparametric Methods," Chapter 42 in *Handbook of Econometrics,* Vol. IV, edited by R.F. Engel and D.L. McFadden, Elsevier.

MATZKIN, R.L. (1999) "Nonparametric Estimation of Nonadditive Random Functions," mimeo, Northwestern University, Invited Lecture at the session on New Developments in the Estimation of Preferences and Production Functions, Cancun, Mexico, August 1999.

MATZKIN, R.L. (2003) "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339-1375.

MATZKIN, R.L. (2004) "Unobservable Instruments," mimeo, Northwestern University.

MATZKIN, R.L. (2005) "Identification in Nonparametric Simultaneous Equations," mimeo, Northwestern University.

MATZKIN, R.L. (2006) "Nonparametric Identification," in *Handbook of Econometrics,* Vol 6, edited by J.J. Heckman and E.E. Leamer, Elsevier, *forthcoming.*

MATZKIN, R.L. and M.K. RICHTER (1991) "Testing Strictly Concave Rationality," *Journal of Economic Theory,* 53, 287-303.

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.

McFADDEN, D.L. (1978) "Cost, Revenue, and Profit Functions," in *Production Economics: A Dual Approach to Theory and Applications, Vol. I: The Theory of Production,* by M. Fuss and D.L. McFadden, eds., North-Holland, Amsterdam.

McFADDEN, D.L. (1981) "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, by C.F. Manski and D.L. McFadden (eds.), MIT Press, Cambridge, MA, pp. 2-50.

McFADDEN, D.L. (1985) Presidential Address, World Congress of the Econometric Society.

McFADDEN, D.L. (2005) "Revealed Stochastic Preferences: A Synthesis," *Economic Theory.*

McFADDEN, D.L. and M.K. RICHTER (1970) "Revealed Stochastic Preferences," mimeo.

McFADDEN, D.L. and M.K. RICHTER (1990) "Stochastic Rationality and Revealed Stochastic Preference," in J. Chipman, D. McFadden, and M.K. Richter (eds.) *Preferences, Uncertainty, and Rationality*, Westview Press, 187-202.

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

NEWEY, W.K. and J.L. POWELL (2003): "Instrumental Variables Estimation for Nonparametric Models," *Econometrica*, 71, 1557-1569..

NEWEY, W.K., J.L. POWELL, and F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica* 67, 565-603.

NG S. and J. PINKSE (1995), "Nonparametric Two-Step Regresion Estimation when Regressors and Errors and Dependent," mimeo, University of Montreal.

PINKSE, J. (2000): "Nonparametric Two-Step Regression Estimation when Regressors and Errors are Dependent," *Canadian Journal of Statistics,* 28-2, 289-300.

POWELL, J.L., J.H. STOCK, and T.M. STOKER (1989) "Semiparametric Estimation of Index Coefficients," *Econometrica,* 57, 1403-1430.

RICHTER, M.K. (1966) "Revealed Prefernce Theory," *Econometrica,* Vol. 34, No. 3, pp 635-645.

RIVERS, D. and Q.H. VUONG (1988): "Limited Information Estimators and Exogneity Tests for Simultaneous Probit Models," *Journal of Econometrics,* 39, pp. 347-366.

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

SAMUELSON, P.A. (1938) "A Note on the Pure Theory of Consumer Behavior," *Economica,* 5, 61-71.

YATCHEW, A. (1985) "A Note on Nonparametric Tests of Consumer Behavior," *Economic Letters,* 18, 45-48.