

NONPARAMETRIC IDENTIFICATION IN STRUCTURAL ECONOMIC MODELS

Rosa L. Matzkin

Department of Economics

University of California - Los Angeles

February 2013

Abstract

Structural economic models allow one to analyze counter-factuals when economic system change and to evaluate the well being of economic agents. A key element in such analysis is the ability to identify the primitive functions and distributions of the economic models that are employed to describe the economic phenomena under study. Recent developments in nonparametric identification of structural models have provided ways to achieve identification of these primitive functions and distributions without imposing parametric restrictions. In this article, I consider a small set of stylized models, and provide insight into some of the approaches that have been taken to develop nonparametric identification results for those models.

1. Introduction

Among the most important objectives of economics are the understanding of how economic systems work, the evaluation of the well being of economic agents, and the prediction of values that economic variables may achieve. Structural econometrics is essential to meet these objectives. The tools structural econometrics develops employ data to test economic models, to measure the preferences of economic agents, and to predict outcomes, specially those outcomes resulting from changes in the economic systems. Nonparametric econometric methods aim at developing such tools without specifying ad-hoc parametric forms for the functions and distributions in the models.

To provide some examples, suppose that a firm is considering to introduce a new product. To predict the profit that such action would generate, it would be necessary to obtain an estimate of the demand for the new product. However, since the product has never been sold, no past observations on the demand for the product would be available to calculate such estimates. Structural econometrics provides tools to predict such demand. As developed in McFadden (1974), using data on the demand for existing products one can estimate the preferences of consumers for the characteristics of the new and its competing products. These preferences can then be used to predict the demands for both the new and the old products when all are available, after the introduction of the new product. McFadden (1974) represented the preferences of a typical consumer for each product as the sum of a linear function of observable characteristics of the product and the consumer and an unobservable random term. The unobservable random term, representing the preferences of the consumer for unobserved characteristics had a known distribution. A large literature has evolved since then, developing results that do not require parametric specifications either for the function of the observable attributes or for the distribution of the unobservable random terms.

When evaluating what segments in a population would be better off and what segments would be worse off as a result of a policy, it is necessary to have knowledge of the distribution of preferences in the population. Structural econometrics can recover such preferences from the distribution of past choices, as in Heckman and Willis (1977), who developed a model of labor participation choice by heterogeneous agents. While their model specified all functions and distributions parametrically, more recent work has considered similar models without specifying parametric assumptions.

As the previous examples illustrate, the tools developed by structural econometrics allow one to predict and evaluate new situations when *some* of the elements in an economic system change, while other stay stable. Key ingredients for this analysis are economic theory models. They specify the connection between the underlying functions one would like to recover and the behavior of the observable variables. In the first example, the underlying function would be that representing the preferences of the typical consumer on the observable characteristics; the observable function would be the demand for existent products. Economic theory models also specify how the stable functions are used to evaluate counter-factuals.

The dependence of structural econometrics on economic theory models is its strength, but it is also its weakness. If the specified models are incorrect, the conclusions obtained by employing those models will in general also be incorrect. Not imposing parametric structures on the functions and distributions in the models reduces the risk of misspecification. Nonparametric tests can be performed to test the validity of models. Nevertheless, misspecification of the models used in the analysis is an issue of concern. Some researchers have responded to these concerns by abandoning the structural approach and focusing on analyses that require no foundation on economic theory models. When the objective of the analysis is prediction when no significant change in the system occurs, this reduced form analysis can be effective. When *all* elements in the economic system stay stable, analyzing the behavior of observable variables does not require reliance on a model. However, when the objective is to estimate the value of a function that is not directly observed, or to predict outcomes when some important elements of an economic system change, understanding the deep behavior of economic variables becomes usually essential. At the very least, specifying the economic theory model on the basis of which conclusions are derived allows one to determine how the conclusions would change if the assumptions did not hold.

The need of relying on economic theory models for the development of econometric methods was recognized since the early times of the Cowles Commission. As Malinvaud (1988) vividly describes, "during the first years a good deal of attention was devoted to direct measurement Cowles Commission Monograph No. 3, ... first published on 1938, ... was entirely devoted to working out monthly indexes of stock prices and annual indexes of stock yields." The motto of the Cowles Commission at the time was "Science is measurement." However, "By 1952 the role of theory become so important that the motto "Science is measurement" no longer seemed appropriate." The suggestion to changing the motto to "Theory and measurement," was immediately accepted. Cowles Commission Monograph No. 10 (1950) contained papers of, quoting again Malinvaud (1988), "the most influential conference on statistical inference in economics ever held." The monograph included the works by R.L. Anderson, T. Haavelmo, H. Hotelling, L. Hurwicz, L.R. Klein, T.C. Koopmans, R. Leipni, H.B. Mann, J. Marschak, G. Tintner, and A. Wald, which built the simultaneous-equation methodology.

Economic theory has been concerned since its early times with the concept of recoverability of primitive functions and with restrictions implied by economic models. Those results have rarely been built upon restrictions requiring that the unknown functions in the models were known up to the value of a finite dimensional parameter. For example, Samuelson (1938) Weak Axiom of Revealed Preference is a set of restrictions on observable choices and budget sets that are necessarily satisfied by any two choices that are generated by a consumer maximizing a common utility function. As shown in Richter (1966), Houthakker (1950)'s Strong Axiom of Revealed Preference characterizes choice behavior generated by the optimization of a common preference over general choice sets. Only weak properties, such as monotonicity, continuity, and convexity of the preferences generating the choices are usually assumed when recovering the set of preferences and utility functions that are consistent with observed choice behavior.

In contrast to the developments in economic theory, econometric methods often started out by specifying functions and distributions up to the value of such a finite dimensional parameter.

Even though the concept of nonparametric identification was introduced very early on by Hurwicz (1950), its value seem to not have been recognized by econometricians till much later.

Econometricians' strong interest in nonparametric estimators seems to have started around the late 70's and 80's. Manski (1975)'s estimation method for a discrete choice model with no parametric assumptions for the distribution of the random term challenged the widely held view that such relaxation of parametric assumptions was impossible. The advances in computer technology possibly had some effect in the development of nonparametric methods, by significantly reducing the cost of applying them. Nonparametric estimation methods are well known to be computational intensive and to be most appropriate for large data sets. There was also increasing evidence during those years that parametric models performed poorly. Gallant (1981,1982) introduced the Fourier Flexible Form for estimation of functions motivating the newly developed estimators as converging in a Sobolev norm, which includes the sup norm of the values and the derivatives of the true function, to a nonparametric function. The flexible parametric forms that were used prior to that development approximated any function up to its second derivative only locally around a point. These second order approximations had been shown to provide very poor global approximations to several functions. Another work that provided evidence of the poor performance of parametric methods was presented by Heckman and Singer (1984). They considered a nonlinear duration model with unobserved heterogeneity, and showed that the parameter of interest can be severely biased due to a misspecification of the distribution of heterogeneity. Heckman and Singer (1984) provided conditions under which the parameter of interest as well as a nonparametric distribution of the unobserved heterogeneity are identified, and developed estimators for both.

Since a nonparametric estimator for a function cannot be shown to be consistent unless the function is identified, the surging interest in nonparametric estimators fueled the study of nonparametric identification. However, the analysis of nonparametric identification has much more value than as a necessary first step in a proof of consistency of an estimator.

As Hurwicz (1950) stated in the early days of the Cowles Commission, "nonparametric identification is a property of a model." One can ask, for example, whether the utility function generating a given consumer demand function is identified (see, e.g. Mas-Colell (1977)), or whether the individual demands of consumers are identified when only the sum of their demand functions is observed. The answer depends on the mapping, which is implied by the model, from the unobserved functions to the observed functions. If the unobserved functions are identified in this nonparametric sense, one can then be more confident about the parametric identification of those same functions. If the unobserved functions are not nonparametrically identified, but they are identified when one imposes on them an ad-hoc parametric specification, then identification is imposed by the ad-hoc restriction, and the conclusions derived from the estimates of such parametric models will strongly depend on those ad-hoc specifications being correct.

When a function in a model is not nonparametrically identified, one can consider several additional restrictions in the model and analyze identification under those restrictions. Often, economic theory implies shape restrictions, such as monotonicity, concavity and homogeneity of degree one, on functions that may not be observed. Those shape restrictions can be employed to restrict the set of functions and achieve identification of otherwise nonidentified models, as shown for example in Matzkin (1992). (See Matzkin (1994) for a review.) An alternative route is to not consider additional restrictions and stop at the smallest set of functions that contains the function of interest. See Manski (1989, 2003) for introductions to this *partial identification* approach. The review article by Tamer (2011) documents earlier and more recent developments on this approach, which analyzes conditions under which a parameter is *set identified*. In the sections that follow, I will review results on *point identification*, which for short, will be called *identification*.

Recently, the shape restrictions that can be derived from economic theory models were used to identify not only unobserved functions but also the values of unobserved variables within the

models. Variables such as the taste of a consumer for a particular product, or the productivity of a worker in a particular job, are essential for the analysis of, correspondingly, models of consumer demand or of labor income. Economic theory models often imply restrictions on the way in which these unobserved variables affect the values of the observed variables. These restrictions have been used recently to develop identification results for nonparametric models where these unobserved variables enter in nonadditive ways in the models. This approach contrast the one where the original economic model is transformed into another one where the unobservables enter in an additive way. When confronting an economic model with data, the *nonparametric nonseparable* approach leaves the model intact, labeling variables as observed or unobserved, depending on the data at hand. In contrast, the *nonparametric separable* approach usually transforms the model into a different one, separating it into functions whose arguments are only observable variables, and functions whose arguments involve the unobservable variables. The analysis of the latter approach usually focuses on the functions of the observable variables. The analysis of the nonparametric nonseparable approach focuses on the analysis of the original functions in the model, whose arguments are the observable and unobservable variables. Hurwicz (1950) noted the importance of allowing nonadditive random terms within econometric models but seemed to recognize that the techniques that were being developed at that time were not amenable to handle such models. Much later, the approach was developed making use of parametric assumptions on all functions and distributions. (See the early works by McFadden (1974), Heckman (1974), and Lancaster (1979).) I will review the nonparametric results that were developed more recently.

In the following sections, I review some of the identification results that have been developed for separable and nonseparable nonparametric models. To avoid blurring the main insights, I analyze the main issues on the most possible stylized models and do not detail technical assumptions. In Section 2, I review the definition of nonparametric identification. I then proceed to analyze three types of models. In Section 3, I deal with models with no endogenous explanatory variables. Models with no simultaneity are analyzed in Section 4. Section 5 contains identification results for models with simultaneity.

2. Definition of nonparametric identification

We follow Hurwicz (1950) by defining a structure as a system of equations and a distribution of exogenous variables, which determine a distribution of endogenous variables. Hurwicz (1950) defined as a *model* a set of such structures with the property that the unknown distributions and unknown functions satisfy a set of restrictions. Two structures are said to be *observationally equivalent* if they generate the same distribution of the observable variables. A feature of a structure, such as a particular function or the value of a particular function at a particular point in its domain, is said to be *identified* if among the set of observationally equivalent admissible structures, the value of the feature does not vary. These definitions make it clear that, as Hurwicz (1950) pointed out, identification is a property that "a model may or may not possess" and that "does not require a parametric representation" for the structures and models.

The analysis of identification is separate from statistical issues, which are dependent on sample size. Identification analysis assumes that the whole probability distribution of the observable variables, rather than a sample from it, is available. When a sample is available, several methods exist to calculate nonparametric estimators for the distribution of the observable variables that, under corresponding assumptions, converge in probability or almost surely to the true function. The analysis of identification starts from the ideal situation where the distribution of the observable variables is known, and analyzes the invertibility properties of the probabilistic model that generates such distribution of observable variables. The focus is typically on features of the distribution of the observable variables or on features of more primitive functions than the distribution of the

observable variables.

When one can find a known, closed form functional defined over the set of possible distributions of the observable variables, such that when evaluated at the true distribution of the observable variables delivers uniquely the feature of interest, it is said that identification is *constructive*. Constructive identification methods indicate in a transparent way the connection between the feature of interest and the distribution of observable variables. They provide a way to read off the distribution of the observable variables the feature of interest. Methods for constructive identification directly lead to methods of estimation for the object of interest. Instead of evaluating the known functional at the true distribution of the observable variables, one evaluates it at a nonparametric estimator for the distribution of the observable variables. The result is a nonparametric estimator for the feature of interest. When the functional is continuous on the set of distribution functions, it is often possible to show that such resulting estimator is consistent. The conditional expectation of one of the observable variables given the other observable variables is an example of a feature that can be identified constructively in an obvious way, by calculating it from the distribution of observable variables using the definition of conditional expectation. Showing constructive identification for features of more primitive functions is often not so obvious. The latter has been the objective of a large literature.

At the heart of the concept of identification is a system of functional equations connecting the known distribution of the observable variables with the unknown functions and distributions generating that known distribution. Identification of one of the unknown elements in the system is analogous to asking whether the system has a unique solution for such unknown element.

To formalize the concepts related to identification, we let X denote the vector of observable variables determined outside the system, referred to as the vector of observable exogenous variables or observable inputs. The marginal distribution of X is known. We let ε denote the vector of unobservable variables determined outside the system, referred to as the vector of unobservable exogenous variables or unobservable inputs. The true distribution of the unobservable and observable exogenous variables will be denoted by $F_{\varepsilon,X}^*$. The vector Y will denote the vector of observable variables determined by the system, referred to as the vector of observable endogenous variables or observable outputs. The system of equations, which specify how the value of Y is determined by the values of X and ε , will be denoted by a vector of known functions, s , which depend on a vector of unknown functions, m^*

$$s(Y, X, \varepsilon; m^*) = 0$$

The set of functions m that satisfy the restrictions that m^* is known to satisfy will be denoted by M . The set of distributions $F_{\varepsilon,X}$ that satisfy the restrictions that $F_{\varepsilon,X}^*$ is assumed to satisfy will be denoted by Γ . One of such restrictions is that all the distributions in Γ share the same known marginal distribution of X , F_X . The system is such that given any function m in M and a distribution $F_{\varepsilon,X}$ in Γ , $s(\cdot; m)$ determines a distribution $F_{Y,X}(\cdot, \cdot; m, F_{\varepsilon,X})$.

Two pairs $(m, F_{\varepsilon,X})$ and $(m^*, F_{\varepsilon,X}^*)$ are *observationally equivalent* if the distributions, $F_{Y,X}(\cdot, \cdot; m, F_{\varepsilon,X})$ and $F_{Y,X}(\cdot, \cdot; m^*, F_{\varepsilon,X}^*)$, generated respectively by $(m, F_{\varepsilon,X})$ and $(m^*, F_{\varepsilon,X}^*)$, are equal. The pair $(m^*, F_{\varepsilon,X}^*)$ is *identified* if no other pair $(m, F_{\varepsilon,X})$ that satisfies the same restrictions that $(m^*, F_{\varepsilon,X}^*)$ has been specified to satisfy is observationally equivalent to $(m^*, F_{\varepsilon,X}^*)$. We next formalize these definitions.

Definition: $(m^*, F_{\varepsilon,X}^*) \in (M \times \Gamma)$ is *identified* in $(M \times \Gamma)$ if for all $(m, F_{\varepsilon,X}) \in (M \times \Gamma)$

such that $(m, F_{\varepsilon, X}) \neq (m^*, F_{\varepsilon, X}^*)$,

$$F_{Y, X}(\cdot, \cdot; m, F_{\varepsilon, X}) \neq F_{Y, X}(\cdot, \cdot; m^*, F_{\varepsilon, X}^*).$$

Given a feature of interest, denote by $\psi(\cdot, \cdot)$ the mapping from $(M \times \Gamma)$ to the set of all possible values, Ω , that the feature of interest may attain for any $(m, F_{\varepsilon, X}) \in (M \times \Gamma)$. Denote by ψ^* the true value of $\psi(m^*, F_{\varepsilon, X}^*)$.

Definition: The feature $\psi^* = \psi(m^*, F_{\varepsilon, X}^*)$ of $(m^*, F_{\varepsilon, X}^*)$ is *identified* if for any $(m, F_{\varepsilon, X}) \in (M \times \Gamma)$ that is observationally equivalent to $(m^*, F_{\varepsilon, X}^*)$,

$$\psi(m, F_{\varepsilon, X}) = \psi(m^*, F_{\varepsilon, X}^*).$$

Constructive identification of a feature $\psi(m^*, F_{\varepsilon, X}^*)$ is defined by the existence of a known functional on the set of distribution functions of the observable variables. Denote by F the set of all distribution functions $F_{Y, X}$ that could be generated by a pair $(m, F_{\varepsilon, X}) \in (M \times \Gamma)$. Denote $F_{Y, X}(\cdot, \cdot; m^*, F_{\varepsilon, X}^*)$, the distribution of the observable variables generated by $(m^*, F_{\varepsilon, X}^*)$, by $F_{Y, X}^*$. Suppose that Φ is a known functional that assigns to each element of F an element of Ω .

Definition: The feature $\psi^* = \psi(m^*, F_{\varepsilon, X}^*)$ of $(m^*, F_{\varepsilon, X}^*)$ is *identified constructively* through the known functional Φ if

$$\psi(m^*, F_{\varepsilon, X}^*) = \Phi(F_{Y, X}^*).$$

As mentioned above, constructive identification leads naturally to the definition of a nonparametric estimator. Denote a nonparametric estimator of $F_{Y, X}^*$ by $\hat{F}_{Y, X}$. If $\psi^* = \psi(m^*, F_{\varepsilon, X}^*)$ is identified constructively through the known functional Φ , then an estimator for it can then be defined by

$$\hat{\psi} = \Phi(\hat{F}_{Y, X})$$

Usually, if $\hat{F}_{Y, X}$ converges in probability to F^* and Φ is continuous, it is possible under additional technical conditions to show that $\hat{\psi}$ converges in probability to ψ^* . Except for Section 5.2, most of the identification results that I will review will be constructive.

3. Non-endogenous explanatory variables

In many econometric models, the explanatory variables are determined by the decisions of individual agents, or by interactions among agents. For example, the years of education of an individual is typically used as an explanatory variable for the individual's wage. Market equilibrium prices are employed to estimate the effect of price changes in consumer demand. When the values of explanatory observable variables are determined by unobservable explanatory variables in the model, identifying the partial effect of a change in the value of the observable explanatory variable, when the values of the unobservable explanatory variables stay constant, may not be possible. To see the complications that can arise in that situation, consider the simple model

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

where m is an unknown function, Y and X are observed, ε is unobserved, and X is an unknown function of ε . Denote the inverse of the later function as $\varepsilon = r(X, \delta)$, where δ is distributed independently of X and unobservable. The distribution of (Y, X) is generated then by the system

$$Y = m(X, r(X, \delta)) = p(X, \delta)$$

for some function p .

Identification of the effect in the value of Y of a change in the value of X when the value of ε stays fixed requires identification of m . However, the distribution of the observable variables (Y, X) is generated by the function p . Identification of m requires additional information about the system allowing to disentangle the effect on Y of X from the effect on Y of ε . We review some of the existing approaches to dealing with such situations in Sections 4 and 5. In this section, I consider models where such problem does not exist. In the models we next consider, the value of a functional of the distribution of the unobservable explanatory variables given the observable explanatory variables is assumed to be fixed across the different values that the observable explanatory variable attains. The constancy of the value of such functional allows to identify the effect of X on Y when the value of ε is fixed.

3.1. Separable models

In separable models, the effect on the endogenous variables of the observable and unobservable explanatory variables is separated into a function that depends only on the observable variables and another function that depends on the unobservable and observable variables. The main object of interest in these models has been usually the function that depends on the observable variables. A simple version of a nonparametric separable model is the additive model

$$(3.1) \quad Y = m(X) + \varepsilon$$

where m is an unknown continuous function, Y is a scalar, X is a vector, both Y and X are observed, and ε is an unobserved random variable. In this model, the effect on Y of a change in the value of X is allowed to be different across different values of X . In contrast, when the function m is specified as a linear function: $m(X) = \alpha + \beta'X$, the effect of X on Y is constant across all values of X .

When one is interested in the effect of X on Y at a particular value, x^* , of X , imposing a linear or other parametric specification on the function m may generate severely biased results. A parametric specification for m imposes a particular global shape on the function m , which may impose local behavior that is very distant from the true behavior of the function. (See, for example, the several examples documented in Hardle (1990), where local peaks of functions were discovered empirically only once nonparametric estimation methods were used.)

A parametric specification on m also hides issues that arise due to support conditions of the probability density of X . A parametric model interpolates between the values of m at the support points of X , identifying the value of m over any upper set of the support of X . The nonparametric approach allows one to clearly distinguish between the conditions necessary to identify m on the support of X and the stronger set of conditions necessary to identify m on a set larger than the support of X . For example, when the support of X is a discrete set of points, no particular shape restrictions will usually be needed to identify nonparametrically the values of m on that discrete set. Nonparametric identification of the values of m on points outside such support would require additional shape restrictions such as, for example, monotonicity, concavity, or homogeneity of degree. Even with additional restrictions, the value of m will typically be only partially identified outside the support of X .

The separable additive model generates the following functional equation in the unknown function m and the unknown conditional distribution $F_{\varepsilon|X}$ of the unobservable random term ε given X :

$$(3.2) \quad F_{Y|X=x}(y) = F_{\varepsilon|X=x}(y - m(x))$$

where $F_{Y|X}$ denotes the known conditional distribution of Y given X . The equation follows, sequentially, by the definition of the conditional distribution of Y given $X = x$, the definition of Y , and the definition of the conditional distribution of ε given $X = x$,

$$\begin{aligned} F_{Y|X=x}(y) &= \Pr(Y \leq y | X = x) \\ &= \Pr(m(X) + \varepsilon \leq y | X = x) \\ &= \Pr(\varepsilon \leq y - m(x) | X = x) \\ &= F_{\varepsilon|X=x}(y - m(x)) \end{aligned}$$

Without additional restrictions, the functional equation (3.2), which contains one known function and two unknown functions, does not possess a unique solution $(m, F_{\varepsilon|X})$. For example, if we take any $\alpha \neq 0$, and define m' and $F'_{\varepsilon|X}$ as

$$m'(x) = m(x) + \alpha \quad F'_{\varepsilon|X=x}(e) = F_{\varepsilon|X=x}(e + \alpha)$$

we can verify that $(m', F'_{\varepsilon|X})$ also satisfies

$$F_{Y|X=x}(y) = F'_{\varepsilon|X=x}(y - m'(x))$$

Nonparametric identification of m requires an additional condition. Usually, the condition takes the form of a location restriction on the distribution of ε given X , such as the condition that a.s in x , the conditional expectation of ε given $X = x$ is zero; i.e., $E(\varepsilon|X = x) = 0$. A median independence restriction, specifying that a.s. in x , $Med(\varepsilon|X = x) = 0$ is often also used as a location restriction. Note that if m is identified, the functional equation becomes an equation in only the unknown function $F_{\varepsilon|X}$. This function can be identified once m is identified by the equation

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

With the mean independence restriction, $m(x)$ is identified as the conditional expectation of Y given $X = x$. Hence, when m is continuous, the function m is identified on the support of X . An analogous argument shows that m is identified on the support of X from the conditional distribution of Y given $X = x$ when the median independence location restriction is adopted, since if $Med(\varepsilon|X = x) = 0$ and $Y = m(X) + \varepsilon$ implies that $Med(Y|X = x) = m(x) + Med(\varepsilon|X = x) = m(x)$.

3.2. Nonseparable models

The separable approach has typically focused on the function of the observable explanatory variables. The terms "noise" or "disturbance", which has usually been used to denote the added unobservable, ε , has given the impression that these were unimportant variables. And in fact, in some models, they were added at the end as an after-thought. The analysis of these added unobservable mostly focused on asking whether they may represent an excluded variable that is correlated with

the vector of observable explanatory variables, X , in which case a constant location restriction may not hold, or whether the variances of $F_{\varepsilon|X}$ was different across different values of X .

In contrast to the separable approach, in the nonseparable approach the focus is on the interactions between the observable and unobservable variables. The object of interest is the original economic model rather than a function that only relates the observable variables. When confronting the original economic model with data, the nonseparable approach uses the data to determine which of the variables in the original model are observable and which are unobservable, but aims at keeping the model intact.

The identification analysis in the nonseparable approach, which proceeds by classifying the variables in the original model as observable or unobservable and analyzing identification of such original model, raises new issues. The number of unobservable random variables in the model may be very large, and the unobservables often enter in the model in nonlinear, nonadditive ways. On the other hand, because these unobservable variables correspond to economic variables in the model, one can exploit the economic model and the implications of economic theory to learn about the restrictions that the values of these unobservable variables must satisfy and the way in which these variables interact with the other variables in the model. This is currently an area of active research.

One of the simplest nonseparable nonparametric models one can consider is

$$(3.3) \quad Y = m(X, \varepsilon)$$

where the scalar Y and the vector X are observed and where the scalar ε is unobserved. An example of such model would be the production function of a typical worker, where Y denotes output, X denotes exogenously determined hours of work, and ε denotes the unobserved productivity of the worker. The unknown function m is then a production function with observable and unobservable inputs. The model allows one to analyze the substitution between hours of work and productivity. The model allows one also to analyze the effect on output of an exogenous change in the amount of hours of work for each individual worker, when each worker is characterized by his or her ability ε .

Model (3.3) is consistent with the following transformed model,

$$(3.4) \quad Y = s(X) + \eta$$

where $s(x)$ is the expected output for workers that work x hours. For any worker working x hours and producing output y , $\eta = y - s(x)$ is the deviation of that worker's output from the expected output. In other words,

$$s(x) = E(Y|X = x) = \int m(x, e) dF_{\varepsilon}(e)$$

and

$$\eta = m(x, e) - \int m(x, e) dF_{\varepsilon}(e).$$

Model (3.4), characterized by s and η , does not provide information on how hours of work and productivity jointly determine output. The function s can be used to determine the change in the expected output when hours of work change, but it cannot be used to determine the change in the output of a particular worker when the hours of work of that particular worker change. The function m in (3.3) can provide an answer to the latter question, by evaluating the change in its value when ε stays fixed and x changes.

Matzkin (1999, 2003) considered identification of the nonseparable model above assuming that m is strictly increasing in ε and that X and ε are independently distributed. In such case, the

functional equation that relates the known distribution of the observable variables to the unknown functions in the model is, for all e and x

$$(3.5) \quad F_\varepsilon(e) = F_{Y|X=x}(m(x, e))$$

where F_ε is the cumulative distribution of ε and $F_{Y|X=x}$ is the conditionally cumulative distribution of Y given $X = x$. This equation follows by the following equalities:

$$\begin{aligned} F_\varepsilon(e) &= \Pr(\varepsilon \leq e) = \Pr(\varepsilon \leq e | X = x) \\ &= \Pr(m(X, \varepsilon) \leq m(x, e) \mid X = x) \\ &= \Pr(Y \leq m(x, e) \mid X = x) = F_{Y|X=x}(m(x, e)) \end{aligned}$$

The first equality follows by the definition of the cumulative distribution function F_ε ; the second equality follows because ε is distributed independently of X ; the third equality follows because m is strictly increasing in ε and therefore, conditional on $X = x$, the event $(\varepsilon \leq e)$ is equivalent to the event $(m(X, \varepsilon) \leq m(x, e))$. The fourth and fifth equality follow, consecutively, by the definition of Y and by the definition of the conditional cumulative distribution function of Y given $X = x$.

Equation (3.5) has several implications. Assume for simplicity that the distribution of ε is strictly increasing at $\varepsilon = e$. This and the strict monotonicity of m imply that the conditional distribution of Y is also increasing. Note that the left hand side of (3.5) does not depend on x . Hence, if we fix the value of ε at $\varepsilon = e$, and vary the value of X from x to x' , we get that

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

Hence, the change in the output of a worker characterized with a level of ability e , when the worker is exogenously required to work x' hours instead of x hours is

$$m(x', e) - m(x, e) = F_{Y|X=x'}^{-1}(F_{Y|X=x}(m(x, e))) - m(x, e)$$

While we do not observe e , we do observe y , and, because of the strict monotonicity of output in ability, to each y there correspond a unique value of ability. Since the output for any particular worker is observed, we can then establish that the change in the output of the worker that has the (unobserved but uniquely determined) ability e , which corresponds to the observed output y , is

$$(3.6) \quad y' - y = F_{Y|X=x'}^{-1}(F_{Y|X=x}(y)) - y$$

The change in output due to an infinitesimal change in x is similarly obtained, assuming differentiability, by differentiating (3.5) and solving for $\partial m(x, e) / \partial x$. The result is

$$(3.7) \quad \frac{\partial m(x, e)}{\partial x} = - \left(\frac{\partial F_{Y|X=x}(y)}{\partial y} \right)^{-1} \left(\frac{\partial F_{Y|X=x}(y)}{\partial x} \right)$$

where $y = m(x, e)$. This is the derivative of the conditional quantile of Y given $X = x$, evaluated at $Y = y$.

Additional conditions, such as homogeneity of degree one restrictions or separability assumptions, can be used to identify the value rather than the derivatives of the function m . (See Matzkin (2003).)

3.3. Multidimensional unobservables

The simple nonseparable model described in the previous section is extremely stylized and restrictive. Not only does it assume independence between the observable and unobservable variables, but also it only consider a scalar unobservable variable. In the production example, unobserved productivity may be composed of several types of productivity, each interacting in a different way within the production function. We next review some of the methods that have been developed to introduce multidimensional unobservable variable in the models.

Additional exogenous variables

One approach to incorporating a large number of unobservable variables in nonparametric models has been to employ a vector of observable explanatory variables of dimension at least as large as the number of unobservable variables in the model. Intuitively, if the effect of an unobservable ε_k within a vector of unobservable variables $\varepsilon = (\varepsilon_1, \dots, \varepsilon_K)$ is in some way connected to the value of an observable X_k in a vector of observable variables $X = (X_1, \dots, X_K)$, the conditional distribution of the endogenous variable Y when the value of X_k varies should provide some information about the value of ε_k . One such particular case (see Matzkin (2003, Appendix A) is where

$$(3.8) \quad Y = \sum_{k=1}^K m^k(X_k, \varepsilon_k)$$

To show that each unknown function m^k can be identified from the distribution of Y conditional on X , Matzkin (2003) considers this model with the requirement that for each k there exists a value, \bar{x}_k , such that when $X_k = \bar{x}_k$, the value of m^k is fixed and known; e.g., $m^k(\bar{x}_k, \varepsilon_k) = 0$ for all values of ε_k . When the value of X_k equals \bar{x}_k , the effect of the function m^k is "shut down". Hence, when, except for $k = 1$, each X_k equals \bar{x}_k , the model becomes

$$Y = m^1(X_1, \varepsilon_1),$$

the same simple model (3.3) whose identification was described in the previous section. Assuming that m^1 is strictly increasing in ε_1 , and that $X, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_K$ are mutually independent, the function m^1 and its derivatives can then be identified from the functional equation

$$F_{Y|X_1=x_1, X_2=\bar{x}_2, \dots, X_K=\bar{x}_K}(m^1(x_1, \varepsilon_1)) = F_{\varepsilon_1}(\varepsilon_1)$$

The rationale behind this approach is that even though the value of the observable dependent variable is determined by a large number of unobservable variables, there is some subpopulation where, except for $k = 1$, each X_k equals \bar{x}_k , where the value of the observable dependent variable depends on only one unobservable variable. Matzkin (2005a) employs this method to identify the preferences of consumers from equilibrium prices, when consumer's demands are not observed. Briesch, Chintagunta, and Matzkin (2010) employ a similar approach to identify an extension of McFadden's (1974) model where the utility functions of the attributes are nonparametric and nonadditive in an unobservable random term. Matzkin (2012) shows how the method can be employed in models with simultaneity and limited observability on the dependent variables.

Repeat observations

Evdokimov (2011) proposed using panel data. One of the models he considered was

$$(3.9) \quad \begin{aligned} Y_1 &= m(X_1, \varepsilon) + \eta_1 \\ Y_2 &= m(X_2, \varepsilon) + \eta_2 \end{aligned}$$

where (Y_1, Y_2, X_1, X_2) is observed and $(\varepsilon, \eta_1, \eta_2)$ is unobserved. The main assumptions that he made were that m is strictly increasing in ε , ε is independent of (X_1, X_2) , the distribution of η_1 given X_1 is independent of $(X_2, \varepsilon, \eta_2)$, the distribution of η_2 given X_2 is independent of $(X_1, \varepsilon, \eta_1)$, and the characteristic functions of η_1 and η_2 conditional, respectively, on X_1 and X_2 , are different from zero.

Evdokimov (2010) noted that when $X_1 = X_2$, the model (3.7) can be rewritten as

$$\begin{aligned} Y_1 &= A + \eta_1 \\ Y_2 &= A + \eta_2 \end{aligned}$$

which is a system whose identification was studied by Kotlarski (1967). Kotlarski showed that if A, η_1 and η_2 are mutually independent, each has at least one absolute moment, $E(\eta_1) = 0$, and η_1 and η_2 have nonvanishing characteristic functions, then the distributions of A, η_1 , and η_2 are identified from the distribution of (Y_1, Y_2) . Evdokimov (2010) extended this result to hold conditional on (X_1, X_2) . He then obtained the identification of the distributions of η_1 given X_1 , η_2 given X_2 , and of A given $(X_1, X_2) = (x, x)$. From the latter distribution, one can identify $m(x, \varepsilon)$, as in Section 3.2.

Averages

A different approach to dealing with a large number of unobserved variables is to obtain an average measure. In such case, no monotonicity or other restrictions are needed. On the other hand, averages do not provide information on the behavior of individual agents.

Suppose that instead of ε being a scalar, as in (3.3), ε is a vector, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_G)$. The nonseparable model can then be specified as

$$Y = m(X, \varepsilon_1, \dots, \varepsilon_K)$$

In this case, a particular value, y , of Y would correspond to a set of values, $E(y)$, of the vector ε , instead of to a unique value of ε . Without further restrictions, it is not possible to identify, for each value of ε in the set $E(y)$ the effect on Y of a change in X . Hoderlein and Mammen (2007) showed that the average derivative of m with respect to X over the set $E(y)$ can be identified. Denote the τ -quantile of the conditional distribution of Y given $X = x$ by $Q(\tau, x)$. Suppose that $y = Q(\tau, x)$. They showed that

$$(3.10) \quad E \left[\frac{\partial m(X, \varepsilon_1, \dots, \varepsilon_K)}{\partial X} \mid X = x, Y = Q(\tau, x) \right] = \frac{\partial Q(\tau, x)}{\partial x}.$$

4. Endogenous explanatory variables in recursive systems

In nonparametric models with endogeneity, two types of distinctive systems, recursive and simultaneous, are usually considered. The interaction between two endogenous variables, Y_1 and Y_2 is *recursive* if Y_2 is an explanatory variable for Y_1 but Y_1 is not an explanatory variable for Y_2 . The interaction between Y_1 and Y_2 is *simultaneous* if Y_2 is an explanatory variable for Y_1 and Y_1 is an explanatory variable for Y_2 .

Consider, for example, an interaction model between two agents, such as two firms each setting the price for its product. Denote by Y_1 the observable decision of agent 1 and by Y_2 the observable decision of agent 2. If agent 2 chooses the value of Y_2 first, and given Y_2 , agent 1 chooses Y_1 , then the relationship is recursive. If both agents make their decisions at the same time, each taking as given the other's decision, as in a Nash equilibrium, then the relationship is simultaneous.

In a labor economics example of a recursive model, Y_1 denotes individual lifetime earnings and Y_2 denotes level of education. The value of Y_2 is chosen first, as a function of expected but not of realized Y_1 . The value of Y_1 is determined next, as a function of Y_2 , as well as of other observable and unobservable variables. Imbens and Newey (2009) describe a nonseparable version of this model, following Card (2001). In the simple version of such model, the typical individual's lifetime earnings is a function of the level of education and productivity, ε_1 . While productivity is unobserved to the agent at the time of deciding the length of education, the value of a signal, ε_2 , correlated with productivity, ε_1 , is known to the agent at such time. A measure, X_2 , determining the cost of a unit of education is observed as well. Denote the cost of education by the value of a function $c(y_2, X_2)$. The agent chooses Y_2 to maximize expected lifetime earnings given the signal, ε_2 , and the cost of education:

$$Y_2 \text{ solves } \text{Max}_{y_2} E [m^1(y_2, \varepsilon_1) | \varepsilon_2, X_2] - c(y_2, X_2)$$

The solution is then a function, m^2 , of X_2 and ε_2 . The recursive model becomes

$$\begin{aligned} Y_1 &= m^1(Y_2, \varepsilon_1) \\ Y_2 &= m^2(X_2, \varepsilon_2) \end{aligned}$$

(See also Heckman and Vytlacil (1998) for a similar model.) As noted in Imbens and Newey (2009), the same structure applies to a nonadditive extension of production functions estimation, as in Mundlak (1963). In such case, Y_1 is output and Y_2 is input.

All the results I will describe can be studied conditional on a vector of observable variables, Z . However, to emphasize the endogeneity aspects of the models and not to crowd the notation, I will consider only the stylized versions of the models, without Z .

4.1. Separable models

Nonparametric separable models with recursive endogeneity were analyzed by Newey, Powell and Vella (1999). (See also Ng and Pinkse (1995) and Pinkse (2000).) A simple version of their model is

$$\begin{aligned} (4.1) \quad Y_1 &= m^1(Y_2) + \varepsilon_1 \\ Y_2 &= m^2(X_2) + \varepsilon_2 \end{aligned}$$

They analyzed identification in this model under the assumption that ε_2 is mean independent of X_2 and that, conditional on ε_2 , ε_1 is mean independent of X_2 . That is,

$$E[\varepsilon_1 | \varepsilon_2, X_2] = E[\varepsilon_1 | \varepsilon_2] \quad \text{and} \quad E[\varepsilon_2 | X_2] = 0$$

The main object of interest in this model is the effect of an exogenous change in Y_2 on the value of Y_1 . In the labor example, this would be the effect on lifetime earning of one more unit in the length of education, or the derivative of m^1 with respect to y_2 , when the value of ε_1 is kept fixed. Newey, Powell, and Vella (1999) showed the identification of this object in a recursive manner. First, the function m^2 is identified using the observable vector (Y_2, X_2) and the location restriction $E[\varepsilon_2|X_2] = 0$, as described in Section 3.1. Specifically,

$$m^2(x_2) = E[Y_2|X_2 = x_2]$$

Given any value (y_2, x_2) of the vector (Y_2, X_2) , the value of ε_2 corresponding to (y_2, x_2) is the deviation of y_2 from the conditional expectation of Y_2 when $X_2 = x_2$:

$$\varepsilon_2 = y_2 - m^2(x_2) = y_2 - E[Y_2|X_2 = x_2]$$

Together with the conditional mean independence restriction

$$E[\varepsilon_1|\varepsilon_2, X_2] = E[\varepsilon_1|\varepsilon_2]$$

this implies that

$$E[\varepsilon_1|Y_2 = y_2, X_2 = x_2] = E[\varepsilon_1|\varepsilon_2 = y_2 - m^2(x_2), X_2 = x_2] = E[\varepsilon_1|\varepsilon_2 = y_2 - m^2(x_2)]$$

Since the distribution of ε_1 given ε_2 is unknown, the conditional expectation of ε_1 given ε_2 will be an unknown function. Denote this function by g . Then, given (y_2, x_2) , $E[\varepsilon_1|\varepsilon_2] = g(\varepsilon_2) = g(y_2 - m^2(x_2))$. Conditioning both sides of the equation

$$Y_1 = m^1(Y_2) + \varepsilon_1$$

on $(Y_2, X_2) = (y_2, x_2)$ then gives

$$h(y_2, x_2) = m^1(y_2) + g(y_2 - m^2(x_2))$$

where $h(y_2, x_2) = E[Y_1|Y_2 = y_2, X_2 = x_2]$ is a known function, and $g(y_2 - m^2(x_2)) = E[\varepsilon_1|Y_2 = y_2, X_2 = x_2]$ is the value of the unknown function g at the known value of its argument, $y_2 - m^2(x_2)$. The identification argument followed by Newey, Powell, and Vella (1999) proceeds by taking derivatives of this functional equation with respect to y_2 and x_2 and solving the resulting linear system for the derivative of m^1 with respect to y_2 . Differentiating with respect to y_2 and x_2 gives

$$\begin{aligned} h_{y_2}(y_2, x_2) &= m_{y_2}^1(y_2) + g_{\varepsilon_2}(y_2 - m^2(x_2)) \\ h_{x_2}(y_2, x_2) &= -g_{\varepsilon_2}(y_2 - m^2(x_2)) m_{x_2}^2(x_2) \end{aligned}$$

When X_2 is a scalar, one obtains that:

$$(4.2) \quad m_{y_2}^1(y_2) = h_{y_2}(y_2, x_2) + \frac{h_{x_2}(y_2, x_2)}{m_{x_2}^2(x_2)}$$

When X_2 is a vector of dimension d_{X_2} , they let

$$D(x_2) = [m_{x_2}^2(x_2) \ m_{x_2}^2(x_2)']^{-1} m_{x_2}^2(x_2)$$

Then, multiplying $h_{x_2}(y_2, x_2)$ by $D(x_2)$ and solving gives

$$(4.3) \quad m_{y_2}^1(y_2) = h_{y_2}(y_2, x_2) + D(x_2) h_{x_2}(y_2, x_2)$$

Hence, the derivative of the function of interest, m^1 , which maps the value of its endogenous argument, y_2 , into y_1 , is identified from the derivatives of the conditional expectation of Y_1 given $Y_2 = y_2$ and $X_2 = x_2$ and derivatives of the conditional expectation of Y_2 given $X_2 = x_2$.

4.2. Nonparametric Nonseparable Models

Control function approach

The nonseparable version of the recursive model is, in its simplest form:

$$(4.4) \quad \begin{aligned} Y_1 &= m^1(Y_2, \varepsilon_1) \\ Y_2 &= m^2(X_2, \varepsilon_2) \end{aligned}$$

with X_2 distributed independently of $(\varepsilon_1, \varepsilon_2)$. In particular, ε_1 is independent of X_2 conditional on ε_2 . If one substitutes the second equation into the first, the resulting system,

$$\begin{aligned} Y_1 &= h^1(X_2, \varepsilon_1, \varepsilon_2) \\ Y_2 &= h^2(X_2, \varepsilon_2) \end{aligned}$$

is triangular in the unobservable variables $(\varepsilon_1, \varepsilon_2)$.

Chesher (2003) and Imbens and Newey (2009) showed identification in this model using a control function approach. (See Heckman and Robb (1985).) The object of interest is the derivative of m^1 with respect to y_2 , when the value of ε_1 is fixed. Given the distribution of the observable vector (Y_1, Y_2, X_2) , one can identify this derivatives in a way parallel to that demonstrated in Section 4.1. The independence of X_2 and $(\varepsilon_1, \varepsilon_2)$ imply that conditional on ε_2 , ε_1 and X_2 are independently distributed. Assuming that m^2 is strictly increasing in ε_2 , it follows that for an unknown function, r^2 , denoting the inverse of m^2 with respect to ε_2 , we have that given (y_2, x_2)

$$\varepsilon_2 = r^2(y_2, x_2)$$

Since ε_1 is independent of X_2 conditional on ε_2 , and, conditional on ε_2 , Y_2 is a function of only X_2 , it follows that ε_1 is independent of Y_2 , conditional on ε_2 . Then, ε_1 is a function of ε_2 and δ , where δ is independent of (Y_2, ε_2)

$$\varepsilon_1 = s(\varepsilon_2, \delta) = s(r^2(y_2, x_2), \delta)$$

Hence,

$$y_1 = m^1(y_2, \varepsilon_1) = m^1(y_2, s(r^2(y_2, x_2), \delta)) = v(y_2, x_2, \delta)$$

Taking derivatives with respect to y_2 and x_2 , one gets

$$\begin{aligned} m_{y_2}^1 + m_{\varepsilon_1}^1 s_{\varepsilon_2} r_{y_2}^2 &= v_{y_2} \\ m_{\varepsilon_1}^1 s_{\varepsilon_2} r_{x_2}^2 &= v_{x_2} \end{aligned}$$

Since v_{y_2} , v_{x_2} , $r_{y_2}^2$, and $r_{x_2}^2$ can be identified, one can solve for $(m_{\varepsilon_1}^1 s_{\varepsilon_1})$ and for $m_{y_2}^1$, getting that

$$m_{y_2}^1 = v_{y_2} - v_{x_2} \frac{r_{y_2}^2}{r_{x_2}^2}$$

Substituting each of the terms by their expressions in terms of the distributions of the observable variables, one gets

$$(4.5) \quad m_{y_2}^1 = - \frac{\left(\frac{\partial F_{Y_1|Y_2=y_2, X_2=x_2}(y_1)}{\partial y_2} \right)}{\left(\frac{\partial F_{Y_1|Y_2=y_2, X_2=x_2}(y_1)}{\partial y_1} \right)} + \frac{\left(\frac{\partial F_{Y_1|Y_2=y_2, X_2=x_2}(y_1)}{\partial x_2} \right)}{\left(\frac{\partial F_{Y_1|Y_2=y_2, X_2=x_2}(y_1)}{\partial y_1} \right)} \frac{\left(\frac{\partial F_{Y_2|X_2=x_2}(y_2)}{\partial y_2} \right)}{\left(\frac{\partial F_{Y_2|X_2=x_2}(y_2)}{\partial x_2} \right)}$$

Hence, the derivative of the function of interest, m^1 , which maps the value of its endogenous argument, y_2 , and ε_1 into y_1 , is identified from the derivatives of the conditional distribution functions of Y_1 given $Y_2 = y_2$ and $X_2 = x_2$, at $Y_1 = y_1$, and the derivatives of the conditional distribution function of Y_2 given $X_2 = x_2$, at $Y_2 = y_2$.

Exchangeability

Altonji and Matzkin (2001, 2005) considered the model

$$Y_1 = m(Y_2, \varepsilon_1)$$

where m is strictly increasing in ε_1 and ε_1 is not distributed independently of Y_2 . In one of their models, they assumed that the distribution of ε_1 is exchangeable in the value of Y_2 and another endogenous variable \tilde{Y}_2 . Formally, their assumption was for any value e of ε and any values t and t'

$$F_{\varepsilon|Y_2=t, \tilde{Y}_2=t'}(e) = F_{\varepsilon|Y_2=t', \tilde{Y}_2=t}(e)$$

An example of an empirical situation where this exchangeability may be satisfied is where Y_1 denotes the educational achievement of a student in a school, Y_2 is the income of his family, and ε_1 includes an unobservable school effect. The school effect would not be independently distributed from the income of the students in the school. However, its distribution could be the same when the income of two families is exchanged. Letting \tilde{Y}_2 denote the income of the family of another student in the class, the above exchangeability condition would be satisfied. Altonji and Matzkin (2001, 2005) employed the normalization that for some specified value \bar{y}_2 , $m(\bar{y}_2, \varepsilon_1) = \varepsilon_1$. This, together with assumptions on the density of ε that implied the strict monotonicity of the conditional distribution of Y_1 given (\bar{Y}_2, \tilde{Y}_2) , was used to show identification of m and of $F_{\varepsilon_1|Y_2}$.

Altonji and Matzkin (2001, 2005)'s main argument was as follows. The strict monotonicity of m in ε and the exchangeability condition implies that for all e, t, t'

$$F_{Y_1|Y_2=t, \tilde{Y}_2=t'}(m(t, e)) = F_{\varepsilon|Y_2=t', \tilde{Y}_2=t}(m(t', e))$$

Since $m(\bar{y}_2, e) = e$, letting $t' = \bar{y}_2$ and $t = y_2$, it follows that

$$F_{Y_1|Y_2=y_2, \tilde{Y}_2=\bar{y}_2}(m(y_2, e)) = F_{\varepsilon|Y_2=\bar{y}_2, \tilde{Y}_2=y_2}(e)$$

The only unknown in this equation is $m(y_2, e)$. Inverting the strictly increasing function $F_{Y_1|Y_2=y_2, \tilde{Y}_2=\bar{y}_2}$, one gets

$$(4.6) \quad m(y_2, e) = F_{Y_1|Y_2=y_2, \tilde{Y}_2=\bar{y}_2}^{-1} \left(F_{\varepsilon|Y_2=\bar{y}_2, \tilde{Y}_2=y_2}(e) \right)$$

This shows that $m(y_2, e)$ is identified. To show that $F_{\varepsilon|Y_2=y_2}(e)$ is identified, they note that the strict monotonicity of m implies that for all y_2 and e

$$F_{\varepsilon|Y_2=y_2}(e) = F_{Y_1|Y_2=y_2}(m(y_2, e))$$

Hence,

$$(4.7) \quad F_{\varepsilon|Y_2=y_2}(e) = F_{Y_1|Y_2=y_2} \left(F_{Y_1|Y_2=y_2, \tilde{Y}_2=\bar{y}_2}^{-1} \left(F_{\varepsilon|Y_2=\bar{y}_2, \tilde{Y}_2=y_2}(e) \right) \right)$$

4.3. Multidimensional unobservables

The models above have the drawback that the dimension of the unobservable variables is very limited. As in the case of Section 3, one can consider also in this case more general models of unobserved heterogeneity at the expense of only identifying average measures, rather than individual measures.

Suppose the object of interest is

$$Y_1 = m(Y_2, \varepsilon_1, \dots, \varepsilon_K)$$

where Y_1 , Y_2 , and an external variable W are observable and where Y_2 and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_K)$ are independent, conditional on W .

Altonji and Matzkin (2001, 2005) considered the *local average response function*, $\beta(y_2)$, defined as the average derivative of m with respect to y_2 over the distribution of ε for those with $Y_2 = y_2$. Formally,

$$\beta(y_2) = \int \frac{\partial m(y_2, \varepsilon_1, \dots, \varepsilon_K)}{\partial y_2} f_{\varepsilon|Y_2=y_2}(\varepsilon_1, \dots, \varepsilon_K) d\varepsilon_1 \cdots d\varepsilon_K$$

Altonji and Matzkin showed that $\beta(y_2)$ can be constructively identified by

$$(4.8) \quad \beta(y_2) = \int \frac{\partial E(Y_1|Y_2=y_2, W=w)}{\partial y_2} f_{W|Y_2=y_2}(w) dw$$

Blundell and Powell (2003) considered the *average structural function*, $G(y_2)$ defined as the average value of the function m when $Y_2 = y_2$, where the average is over the marginal distribution of ε . Formally,

$$G(y_2) = \int m(y_2, \varepsilon_1, \dots, \varepsilon_K) f_{\varepsilon_1, \dots, \varepsilon_K}(\varepsilon_1, \dots, \varepsilon_K) d\varepsilon_1 \cdots d\varepsilon_K$$

Blundell and Powell showed that $G(y_2)$ can be constructively identified by

$$(4.9) \quad G(y_2) = \int E(Y_1 | Y_2 = y_2, W = w) f_W(w) dw.$$

Imbens and Newey (2009) considered the τ -quantile of $m(y_2, \varepsilon_1, \dots, \varepsilon_K)$, when y_2 is fixed, denoted as $q(y_2, \tau)$. They showed that when the support of W conditional on $Y_2 = y_2$ is constant across y_2 , the τ -quantile of $m(y_2, \varepsilon)$ can be constructed from the distribution of (Y_1, Y_2, W) by

$$(4.10) \quad q(y_2, \tau) = G^{-1}(\tau, y_2)$$

where

$$(4.11) \quad G(y_1, y_2) = \int F_{Y_1 | Y_2 = y_2, W = w}(y_1) f_W(w) dw.$$

5. Simultaneity

The most general model with endogeneity is one that allows the values of all endogenous variables to be determined simultaneously. Several identification results for nonparametric models with simultaneity have been developed. One line of research analyzes identification using conditional moment restrictions generated with observable variables that are excluded from the functions of interest. Identification takes the form of asking whether one or a system of integral equations has a unique solution. Uniqueness of the solution has been studied under assumptions on the conditional distribution of the observable endogenous explanatory variable, given the excluded variable.

Another line of research focuses on the whole system of equations. Identification, in this approach, has been studied under assumptions on the unknown functions in the system and on the joint distribution of the observable and unobservable exogenous variables.

I first provide some examples of economic models where identification requires methods that deal with simultaneity. I then review the results that have been developed to deal with simultaneity. As in the previous section, to concentrate on presenting the main identification insights, I consider the simplest versions of the models. The results can be easily extended to models where all the functions depends on a vector of observable variables, Z .

5.1. Examples of models with simultaneity

The textbook example of a model with simultaneity is the *demand and supply model*. The quantity sold in a typical market, Q , and the price of the product in that market, P , are observed, as well as the income of the consumers, I , and the input costs of the producers, W . The objects of

interest are the demand function, $D(P, I, \varepsilon_D)$, and the supply function, $S(Q, W, \varepsilon_S)$. The demand function $D(P, I, \varepsilon_D)$ denotes the quantity, Q , that consumers are willing to buy when the price of the product is P , the consumers' income is I , and ε_D is the value of a vector of unobserved variables of the consumers influencing their demand. The supply function $S(Q, W, \varepsilon_S)$ denotes the price, P , that producers are willing to accept for producing the quantity Q when the input price is W and ε_S is the value of a vector of unobserved variables that influence the price producers are willing to accept. The system that determines the observed Q and P is then

$$\begin{aligned} Q &= D(P, I, \varepsilon_D) \\ P &= S(Q, W, \varepsilon_S) \end{aligned}$$

The observable values of Q and P are the values that satisfy both equations. Assuming that this system has a unique solution for $(I, W, \varepsilon_D, \varepsilon_S)$, the observable distribution of the vector of endogenous variables (Q, P) , conditional on the vector of observable exogenous variables (I, W) is generated by

$$\begin{aligned} Q &= h^1(I, W, \varepsilon_D, \varepsilon_S) \\ P &= h^2(I, W, \varepsilon_D, \varepsilon_S) \end{aligned}$$

for some functions h^1 and h^2 . Unlike the triangular model analyzed in Section 4, in this model all the endogenous variables are determined by the unobservable variables that enter in all the equations in the system.

Multidimensional optimization models also generate models with simultaneity. Suppose each member in a population chooses the value of a vector Y by maximizing a function $V(Y, X, \varepsilon)$ where ε is a vector representing unobservable taste or characteristics of the individual, and X is an observable vector of characteristics of the member or of the constraints the individual faces. The observed value of Y chosen by a typical agent with unobservable ε satisfies

$$Y = \arg \max_y V(y, X, \varepsilon)$$

The observed Y satisfies the First Order Conditions for maximization

$$V_y(Y, X, \varepsilon) = 0.$$

This is a system of simultaneous equations where the vector of unknown function V_y is the gradient of the objective function V . Integration of the gradient yields the object function. Hence, if the system V_y and the distribution of ε are identified, one can identify the distribution of value functions over the population. Blundell, Kristensen, and Matzkin (2009, 2012) analyzed identification of consumer demand models using this approach. In their models, the function V is the utility function after substituting the value of one of the coordinates of Y by the value determined by the budget constraint, and ε is a vector of unobservable taste for the commodities.

In *markets for differentiated products*, one observes the quantity sold of each of a finite number of products and the prices of the products. The quantity demanded of each product depends on the prices, observed characteristics, and unobserved characteristics of all products, generating a system of demands

$$\begin{aligned} Q_1 &= D^1(P_1, \dots, P_G, X_1, \dots, X_G, \varepsilon_1, \dots, \varepsilon_G) \\ &\dots \\ Q_G &= D^G(P_1, \dots, P_G, X_1, \dots, X_G, \varepsilon_1, \dots, \varepsilon_G) \end{aligned}$$

where P_g denotes the market price of product g , X_g denotes the vector of observable characteristics of product g , ε_g denotes the unobserved characteristic of product g , and Q_g denotes the observed quantity demanded of product g . These models, which are extensions of the discrete choice model developed by McFadden (1974), were analyzed parametrically by Berry, Levinshon and Pakes (1995). Recently, Berry and Haile (2009) analyzed the nonparametric identification of this system of equations, using functional restrictions and instruments for (P_1, \dots, P_G) .

5.2. Conditional Moments Approach

5.2.1. A separable model

A simple version of the model considered in the Conditional Moments Approach is the separable model

$$Y_1 = m(Y_2) + \varepsilon$$

where the function m is unknown, (Y_1, Y_2) is a continuously distributed random vector of observable variables, and for some observable X , the conditional expectation of ε given X is a.s. zero, $E(\varepsilon|X) = 0$. Newey and Powell (1989, 2003), Ai and Chen (2003), Darolles, Florens, and Renault (2002), and Hall and Horowitz (2003) first considered this model. Taking conditional expectations, given X , on both sides of this equation results in the moment condition

$$(5.1) \quad E(Y_1|X = x) = \int m(y_2) f_{Y_2|X=x}(y_2) dy_2.$$

In this expression, the conditional expectation, $E(Y_1|X = x)$, of Y_1 given $X = x$, and the conditional density, $f_{Y_2|X=x}$, of Y_2 given $X = x$, can be calculated from the distribution of the observable variables (Y_1, Y_2, X) . Hence, they are identified regardless of whether the function m is or is not identified. After substituting $E(Y_1|X = x)$ and $f_{Y_2|X=x}(y_2)$ by their expressions in terms of the distribution of (Y_1, Y_2, X) , the integral equation (5.1) becomes a functional equation in only one unknown function, m . The function m is identified if (5.1) has a unique solution. Newey and Powell (1989, 2003) showed that a completeness condition on $f_{Y_2|X}$ is necessary and sufficient for the existence of a unique solution m to the integral equation (5.1). The completeness condition is satisfied iff for all functions $\delta(y_2)$ with finite expectation, $E[\delta(y_2)|X = x] = 0$ implies that $\delta(y_2) = 0$. Suppose for example, that the conditional density of Y_2 given X is of the exponential type, i.e., $f_{Y_2|X=x}(y_2) = t(y_2) r(x) \exp(\mu(x)\tau(y_2))$ with $t(y_2) > 0$ and $\tau(y_2)$ 1-1. Then, completeness is satisfied. (See Newey and Powell (2003).)

Das (2004) and Newey and Powell (2003) considered identification of this model when both Y_2 and X are discrete. Denote the support of Y_2 and X by, respectively, $\{y_{2,1}, \dots, y_{2,S}\}$ and $\{x_1, \dots, x_T\}$. Let P denote the $S \times T$ matrix whose ij -th elements is $\Pr(Y_2 = y_{2,i}|X = x_j)$. Their necessary and sufficient condition for identification of m on $\{y_{2,1}, \dots, y_{2,S}\}$, in this case, is that the rank of P is S .

5.2.2. A nonseparable model

When the model is nonseparable, the question of identification using an external variable, X , has also been studied as a uniqueness question of an integral equation. Chernozhukov and Hansen

(2005), Chernozhukov, Imbens, and Newey (2005), and Chen, Chernozhukov, Lee, and Newey (2011) provided identification results for the function m in the nonseparable model

$$(5.2) \quad Y_1 = m(Y_2, \varepsilon_1)$$

where, for all values, y_2 , of Y_2 , m is strictly increasing in the scalar ε_1 . They assumed that ε_1 is independent of an observable X , generating the nonlinear quantile restriction that for all τ

$$\Pr(\varepsilon_1 \leq \tau | X = x) = \tau \quad a.s.$$

The definition of Y_1 and the strict monotonicity of m in ε_1 imply the conditional moment restriction

$$E[1(Y_1 \leq m(Y_2, \tau)) - \tau | X = x] = 0 \quad a.s.$$

For the special case where the support of Y_2 is $\{0, 1\}$ and the support of X is $\{0, 1\}$, Chernozhukov and Hansen (2005, Theorem 2) showed identification under the assumption that the matrices

$$(5.3) \quad \pi'(\tilde{y}_0, \tilde{y}_1) = \begin{bmatrix} f_{Y_1, Y_2 | X=0}(\tilde{y}_0, 0) & f_{Y_1, Y_2 | X=0}(\tilde{y}_1, 1) \\ f_{Y_1, Y_2 | X=1}(\tilde{y}_0, 0) & f_{Y_1, Y_2 | X=1}(\tilde{y}_1, 1) \end{bmatrix}$$

for all $(\tilde{y}_0, \tilde{y}_1)$ in a set that depends on τ are full rank. For the case where Y_2 is continuous, they showed identification under a bounded completeness condition. (See Chernozhukov and Hansen (2005) for details.)

5.3. Simultaneous System Approach

The simultaneous system approach focuses on the system of structural equations and analyzes identification under restrictions on the unknown functions in the system and on the distribution of the exogenous variables. In this approach, the identification analysis exploits the restrictions that the economic model implies on the structural model. Hence, it is more tied up to the structural model under consideration, and the conditions for identification are more primitive, than those that have been used in the Conditional Moments Approach. Another difference is that the Simultaneous System Approach is more focused on the identification of the value of m at particular points in its domain.

In econometrics textbooks, linear simultaneous equations models are typically specified as

$$(5.4) \quad \varepsilon = AY + BX$$

where, following closely the notation in the previous sections, Y is a vector of G observable endogenous variables, X is a vector of K observable exogenous variables, and ε is a vector of G unobservable exogenous variables. The interactions among the endogenous variables is represented in the $G \times G$ matrix A , while the effect of the observable exogenous variables, or inputs, is represented in the $G \times K$ matrix B . If the matrix A is invertible, one can solve for Y , by premultiplying the above equation by the inverse, $(A)^{-1}$, of A , yielding

$$Y = \Pi X + \nu$$

where $\Pi = -(A)^{-1}B$ and $\nu = (A)^{-1}\varepsilon$. The resulting system is the reduced form system. The identification of the true values of the matrices A and B , and of the distribution of ε has been

the object of study in the works by Koopmans (1949), Koopmans, Rubin, and Leipnik (1950), Haavelmo (1943, 1944), and Fisher (1966), among others. (See Hausman (1983) and Hsiao (1983) for an early review.)

The nonparametric version of the model (5.4) is

$$(5.5) \quad \varepsilon = r(Y, X)$$

where r is a vector of unknown functions that map the observable vector (Y, X) into the value of the unobservable vector ε in R^G . The conditions imposed on this model, to analyze its identification, have generalized the conditions used in linear models. In particular, it has been usually assumed that conditional on X , the relationship between Y and ε is 1-1, and that the vectors ε and X are independently distributed. The independence condition may be satisfied conditional on a vector Z . Hence, in many situations it may not be too strong. The 1-1 restriction, however, may be overly strong. If, for example, Y is the vector of equilibrium prices in a general equilibrium model, the restrictions guaranteeing that the relationship between Y and ε is 1-1 are the same as those guaranteeing a unique equilibrium. More generally, the conditions are the same as those guaranteeing invertibility in nonlinear equations. (See Gale and Nikaido (1965) for one set of such conditions.)

Roehrig (1988), following Brown (1983), proposed identification results for the model (5.5) under the two assumptions that, conditional on X , the relationship between Y and ε is 1-1, and that ε and X are distributed independently. Following Benkard and Berry (2006) counterexamples on Roehrig (1988), Matzkin (2008) developed a new set of identification results. Her results were based on the transformation of variables equation

$$(5.6) \quad f_{Y|X=x}(y) = f_{\varepsilon|X=x}(r(y, x)) \left| \frac{\partial r(y, x)}{\partial y} \right|,$$

which maps the function r and the conditional density of ε given $X = x$ to a conditional density $f_{Y|X}$ of the vector of observable endogeneous variables, Y , given $X = x$. Matzkin (2008) assumed differentiability and full support assumptions on all functions and densities, in addition to the 1-1 assumption on the relationship between Y and ε , conditional on any value x for X , and the independence between ε and X . The pair $(f_{\varepsilon, X}, r)$ was defined to be observationally equivalent to another pair $(\tilde{f}_{\varepsilon, X}, \tilde{r})$, when F_X is known, if and only if for all (y, x)

$$(5.7) \quad f_{\varepsilon}(r(y, x)) \left| \frac{\partial r(y, x)}{\partial y} \right| = \tilde{f}_{\varepsilon}(\tilde{r}(y, x)) \left| \frac{\partial \tilde{r}(y, x)}{\partial y} \right|$$

Based on this definition, Matzkin (2008) provided several characterizations of observational equivalence in terms of only \tilde{r} , r , and f_{ε} .

Constructive identification results, where a feature $\mu(r, f_{\varepsilon})$ of (r, f_{ε}) is characterized by a completely specified functional Φ on the distribution of observable variables $F_{Y, X}^*$ were developed in Matzkin (2007, 2010, 2011) and in Berry and Haile (2009, 2011) for particular specifications of the model (5.5). These characterizations have been derived from (5.6). I describe two such constructive identification results in the following sections.

5.3.1. A special case with exclusive regressors

A particularly useful case of the model (5.5) is the structural model, considered in Matzkin (2005), where:

$$\begin{aligned}\varepsilon^1 &= s^1(y) + x_1 \\ \varepsilon^2 &= s^2(y) + x_2 \\ &\dots \\ \varepsilon^G &= s^G(y) + x_G\end{aligned}$$

with ε distributed independently of (X_1, \dots, X_G) and the relationship between ε and y being 1-1. The latter assumption implies that the Jacobian determinant $|\partial s(y) / \partial y|$ is different from zero.

A constructive method to identify the values of $s^1(y), \dots, s^G(y)$ in this model was stated in Matzkin (2007, Section 2.1.4), in the context of a consumer demand model. In that result, it was assumed that f_ε is differentiable and is known to possess a unique mode, whose value is known. Denote this value by ε^* . The method uses the transformation of variables equation. Due to the additivity in x , and the independence between X and ε , the transformation equation (5.6) becomes

$$f_{Y|X}(y) = f_\varepsilon(s(y) + x) \left| \frac{\partial s(y)}{\partial y} \right|$$

The condition on the density f_ε is that

$$\frac{\partial f_\varepsilon(\varepsilon)}{\partial \varepsilon} = 0 \quad \Leftrightarrow \quad \varepsilon = \varepsilon^*$$

and that ε^* is known. Differentiating the transformation of variables equations with respect to x , one gets

$$\frac{\partial f_{Y|X}(y)}{\partial x} = \frac{\partial f_\varepsilon(s(y) + x)}{\partial \varepsilon} \left| \frac{\partial s(y)}{\partial y} \right|$$

Since $|\partial s(y) / \partial y| \neq 0$, the only way in which the derivative of the conditional density of Y given $X = x$ can equal zero is if the derivative of f_ε with respect to ε is evaluated at the mode, ε^* . But this implies that at the value of x for which the derivative of the conditional density of Y given $X = x$ is zero,

$$s(y) + x = \varepsilon^*$$

Hence, given y , for $x(y)$ is such that

$$\frac{\partial f_{Y|X=x(y)}(y)}{\partial x} = 0$$

one has that

$$s(y) = \varepsilon^* - x(y)$$

In summary, the result states that when the density of ε possesses only one mode and the value of the mode is known, one can identify the value of the function s at y , by finding the value $x(y)$

at which the gradient with respect to x of $f_{Y|X=x(y)}(y)$ equals zero. The value of $s(y)$ equals the mode minus $x(y)$.

The identification of $f_\varepsilon(e)$ follows from the identification of (i) $s(y)$, (ii) $f_{Y|X=x}(y)$ when $x = s(y) - e$, and of (iii) the derivatives of $s(y)$. When these three are identified, $f_\varepsilon(e)$ can be calculated using the transformation of variables equation:

$$f_\varepsilon(e) = f_{Y|X=s(y)-e}(y) \left| \frac{\partial s(y)}{\partial y} \right|^{-1}$$

Other constructive identification results for this model were given in Berry and Haile (2009, 2011), in the context of a model of demand for differentiated products with nonparametric demand and supply equations, and in Matzkin (2010, 2011), using average derivatives to exploit the additive linear specification of X .

5.3.2. Special cases of two equations, one-instrument models

Often, the object of interest is a function within a system, such as the function m^1 in the system

$$(5.8) \quad \begin{aligned} Y_1 &= m^1(Y_2, \varepsilon_1) \\ Y_2 &= m^2(Y_1, X_2, \varepsilon_2) \end{aligned}$$

Matzkin (2010, 2011) provided several identification results for the derivative of m^1 with respect to y_2 in this model. She assumed that m^1 is invertible in ε_1 and m^2 is invertible in ε_2 . The system could then be written as

$$(5.9) \quad \begin{aligned} \varepsilon_1 &= r^1(Y_1, Y_2) \\ \varepsilon_2 &= r^2(Y_1, Y_2, X_2) \end{aligned}$$

Matzkin (2010) assumed that $(\varepsilon_1, \varepsilon_2)$ is distributed independently of X_2 and that conditional on X_2 , the relationship between Y and ε is 1-1.

System (5.8) is the extension to models with simultaneity of the recursive system

$$\begin{aligned} Y_1 &= m^1(Y_2, \varepsilon_1) \\ Y_2 &= m^2(X_2, \varepsilon_2) \end{aligned}$$

specified in (4.2). In the recursive system (4.2), Y_1 is not an argument in m^2 . This allows to identify first elements of m^2 that can be employed in the identification of elements of m^1 . In recursive systems, the first step does not require any knowledge of m^1 or of f_{ε_1} . In non-recursive systems, the inclusion of Y_1 as an argument in m^2 does not allow, in general, to identify elements of m^2 separately from m^1 .

If, however, in the simultaneous equations system (5.8), m^1 and m^2 were linear and additive in, respectively, ε_1 and ε_2 , the elements of m^1 could be identified recursively. Suppose

$$\begin{aligned} Y_1 &= \beta Y_2 + \varepsilon_1 \\ Y_2 &= \gamma Y_1 + \alpha X_2 + \varepsilon_2 \end{aligned}$$

where β, γ , and δ are parameters of unknown values satisfying $(1 - \beta\gamma) \neq 0$. Solving this system for Y_2 , gives a function that is linear in X_2 and additive in an unobservable η :

$$Y_2 = \phi X_2 + \eta$$

The system

$$\begin{aligned} Y_1 &= \beta Y_2 + \varepsilon_1 \\ Y_2 &= \phi X_2 + \eta \end{aligned}$$

can be identified using, among others, the methods described in Section 4.

When the functions m^1 and m^2 in (5.8) are not linear in Y , solving (5.8) for Y_2 will result in a function of three arguments, two of which are ε_1 and ε_2 :

$$Y_2 = h^2(X_2, \varepsilon_1, \varepsilon_2)$$

5.3.2.1. Triangular Blundell and Matzkin (2010) characterized the set of systems (5.8) that are observationally equivalent to (4.2). They showed that for both systems to be observationally equivalent a property, which they called "control function separability," has to be satisfied. For (5.9) to satisfy control function separability it must be that for some functions v and q , each invertible in its first coordinate

$$r^2(Y_1, Y_2, X_2) = v(q(Y_2, X), r^1(Y_1, Y_2))$$

This condition allows Y_1 to be an argument of the function m^2 but only as long as Y_1 enters r^2 inside an index function equal to r^1 .

5.3.2.2. Separable Inverse Matzkin (2010, 2011) provided several observational equivalence results and constructive identification results for the identification of the derivatives of m^1 in the system (5.8), without assuming control function separability. To describe one of her results, closely related to the one in Section 5.3.2, assume that the system (5.9) is of the form

$$\begin{aligned} \varepsilon_1 &= s^1(y_1, y_2) \\ \varepsilon_2 &= s^2(y_1, y_2) + x_2 \end{aligned}$$

Let (y_1, y_2) be a specified value of $Y = (Y_1, Y_2)$. Matzkin (2010) assumed that there exists two values ε'_2 and ε''_2 of ε_2 such that

$$\frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon'_2)}{\partial \varepsilon_2} = \frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon''_2)}{\partial \varepsilon_2} = 0$$

and

$$\frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon'_2)}{\partial \varepsilon_1} \neq \frac{\partial \log f_\varepsilon(\varepsilon_1, \varepsilon''_2)}{\partial \varepsilon_1}$$

where $\varepsilon_1 = r^1(y_1, y_2)$. She then showed a two step procedure to constructively identify $\partial m^1(y_2, \varepsilon_1) / \partial y_2$. First, find any values x'_2 and x''_2 satisfying

$$\frac{\partial \log f_{Y|X_2=x'_2}(y)}{\partial x_2} = \frac{\partial \log f_{Y|X_2=x''_2}(y)}{\partial x_2} = 0$$

and

$$\frac{\partial \log f_{Y|X_2=x'_2}(y)}{\partial y_1} \neq \frac{\partial \log f_{Y|X_2=x''_2}(y)}{\partial y_1}$$

Next, read the derivative of m^1 with respect to y_2 off the conditional density of Y given X_2 , when $Y = (y_1, y_2)$ and $X_2 = x'_2, x''_2$, by

$$(5.10) \quad \frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = - \frac{\left[\frac{\partial \log f_{Y|X_2=x'_2}(y)}{\partial y_2} - \frac{\partial \log f_{Y|X_2=x''_2}(y)}{\partial y_2} \right]}{\left[\frac{\partial \log f_{Y|X_2=x'_2}(y)}{\partial y_1} - \frac{\partial \log f_{Y|X_2=x''_2}(y)}{\partial y_1} \right]}$$

Matzkin (2010) provided also constructive identification of the derivative of m^1 employing conditions on the second order derivatives of $\log(f_\varepsilon)$ and corresponding second order derivatives of $\log(f_{Y|X_2})$.

5.3.3. Identification in general models

Identification results in general models can be obtained by applying the observational equivalence results in Matzkin (2008). Taking logs and differentiating (5.7) with respect to y and x , one gets

$$\begin{aligned} \tilde{r}'_y \tilde{q}_\varepsilon + \tilde{\gamma}_y &= r'_y q_\varepsilon + \gamma_y \\ \tilde{r}'_x \tilde{q}_\varepsilon + \tilde{\gamma}_x &= r'_x q_\varepsilon + \gamma_x \end{aligned}$$

where $\tilde{q}_\varepsilon = \partial \log f_\varepsilon(\tilde{r}(y, x)) / \partial \varepsilon$, $q_\varepsilon = \partial \log f_\varepsilon(r(y, x)) / \partial \varepsilon$, \tilde{r}'_y and r'_y denote, respectively, the transpose of the matrices of partial derivatives with respect to y , \tilde{r}'_y and r'_y , \tilde{r}'_x and r'_x denote, respectively, the transpose of the matrices of partial derivatives with respect to x , \tilde{r}'_x and r'_x , and where $\tilde{\gamma}_y, \tilde{\gamma}_x, \gamma_y$, and γ_x denote the derivatives with respect to y and x of $\log(|\tilde{r}_y|)$ and $\log(|r_y|)$. Based on these equations, Matzkin (2008) provided several characterizations of observational equivalence between \tilde{r} and r .

Constructive identification results in general models can be obtained similarly from (5.6). Taking logs and differentiating (5.6), one gets

$$\begin{aligned} g_y &= r'_y q_\varepsilon + \gamma_y \\ g_x &= r'_x q_\varepsilon + \gamma_x \end{aligned}$$

where $g_y = \partial \log(f_{Y|x=x}(y)) / \partial y$ and $g_x = \partial \log(f_{Y|x=x}(y)) / \partial x$ are observable. (See Matzkin (2010, 2013) and Berry and Haile (2011).)

6. Conclusions

I have reviewed some of the main developments on nonparametric identification in structural economic models. I considered models with non-endogenous explanatory variables, with recursive endogeneity, and models with simultaneity. In each case, I have described some of the main insights that have been used to identify separable and nonseparable models. The results have been described on very stylized models. Analyzing nonparametric identification in realistic situations requires combining and extending these results, as well as adding the insights from the large important literature that has been left unmentioned in this article.

Acknowledgments

The support of the NSF through grant SES-1062090 is gratefully acknowledged. I have also benefited from the input of an editor, Richard Blundell, a referee, discussions with James Heckman, and comments of graduate students at UCLA and of participants at a Torcuato Di Tella University seminar.

Disclosure Statement

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

References

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 R.L. MATZKIN (2001): "Panel Data Estimators for Nonseparable Models with Endogenous Regressors", NBER Working Paper T0267. Revision of 1997 working paper.

ALTONJI, J.G. AND R.L. MATZKIN (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 3, 1053-1102.

BENKARD, C.L. and S. BERRY (2004) "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on B. Brown (1983) and Roehrig (1988), Cowles Foundation Discussion Paper #1482.

BERRY, S.T. and P.A. HAILE (2009) "Identification of Discrete Choice Demand from Market-Level Data," mimeo, Yale University.

BERRY, S.T. and P.A. HAILE (2011) "Identification in a Class of Nonparametric Simultaneous Equations Models," Cowles Foundation Discussion Paper No. 1787R.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995). "Automobile Prices in Market Equilibrium," *Econometrica*, vol. 63(4), pP. 841-90.

BLUNDELL, R., D. KRISTENSEN, and R.L. MATZKIN (2011) "Bounding Quantile Demand Functions using Revealed Preference Inequalities", Cemmap Discussion Paper CWP21/11.

BLUNDELL, R., D. KRISTENSEN, and R.L. MATZKIN (2011) "Consumer Demand with Unobserved Heterogeneity," mimeo, <http://www.econ.ucla.edu/faculty/?b=Matzkin>.

BLUNDELL, R. and R.L. MATZKIN (2010) "Conditions for the Existence of Control Functions in Nonseparable Simultaneous Equations Models," Cemmap Discussion Paper CWP28/10.

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 Economic Studies*, 71 (3), 581-913.

BRIESCH, R.A, P.K. CHINTAGUNTA and R.L. MATZKIN.(2010) "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," *Journal of Business and Economic Statistics*, Vol. 28, No. 2, pp. 291-307.

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.

CARD, D. (2001) "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, Vol. 69, 1127-1160.

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

CHERNOZHUKOV, V., IMBENS, G., NEWEY, W. (2007). "Instrumental variable estimation of nonseparable models". *Journal of Econometrics* 139 (1), 4-14.

CHEN, X., CHERNOZHUKOV, S. LEE, and W. NEWEY (2011) "Local Identification of Nonparametric and Semiparametric Models," Cowles Foundation Discussion Paper No. 1795.

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

CHIAPPORI, P-A, and I. KOMUNJER (2009) "On the Nonparametric Identification of Multiple Choice Models," discussion paper, UCSD.

DAROLLES, S., J.P. FLORENS, and E. RENAULT (2000), "Nonparametric Instrumental Regression," mimeo, IDEI, Toulouse.

DAS, M. (2004) "Instrumental Variables Estimators of Nonparametric Models with Discrete Endogenous Regressors," *Journal of Econometrics*, 124, 335-361.

EVDOKIMOV, K. (2011), "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," mimeo, Princeton University.

FISHER, F.M. (1959). "Generalization of the Rank and Order Conditions for Identifiability". *Econometrica* 27 (3), 431-447.

FISHER, F.M. (1961) "Identifiability Criteria in Nonlinear Systems," *Econometrica*, 29, 574-590.

FISHER, F.M. (1965) "Identifiability Criteria in Nonlinear Systems: A Further Note," *Econometrica*, 33, 197-205.

- FISHER, F.M. (1966) *The Identification Problem in Econometrics*. New York: McGraw-Hill.
- FRISCH, R. A.K. (1934) *Statistical Confluence Analysis by Means of Complete Regression Systems*, Publication No. 5. Oslo: Universitets Okonomiske Institutt.
- FRISCH, R. A.K. (1938) "Statistical versus Theoretical Relations in Economic Macrodynamics," Memorandum prepared for a conference on Cambridge, England, July 18-20, 1938, to discuss drafts of Tinbergen's League of Nations publications.
- GALE, D. and H. NIKAIDO (1965), "The Jacobian Matrix and Global Univalence of Mappings," *Math. Annalen* 159, 81-93.
- GALLANT, A.R. (1981), "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form," *Journal of Econometrics*, No. 15, pp. 211-245.
- GALLANT, A.R. (1982), "Unbiased Determination of Production Technologies," *Journal of Econometrics*, 20.
- GOLDBERGER, A.S. (1972). "Structural equation methods in the social sciences". *Econometrica* 40, 979-1001.
- HAAVELMO, T. (1943). "The Statistical Implications of a System of Simultaneous Equations". *Econometrica* 11 (1), 1-12.
- HAAVELMO, T. (1944) "The Probability Approach in Econometrics," *Econometrica*, 12, Supplement, July.
- HALL, P. and J.L. HOROWITZ (2005), "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 33, 2904-2929.
- HARDLE, W. (1991) *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.
- HAUSMAN, J.A. (1983) "Specification and Estimation of Simultaneous Equation Models," in *Handbook of Econometrics*, Vol. I, edited by Z. Griliches and M.D. Intriligator.
- HECKMAN, J.J. (1974) "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42, 679-693.
- HECKMAN, J.J. and R. ROBB (1985) "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, edited by J.J. Heckman and B. Singer, Econometric Society Monograph 10, Cambridge University Press, Cambridge, U.K.
- HECKMAN, J.J. and B. SINGER (1984) "A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, The Identifiability of the Proportional Hazard Model," *Econometrica*, 52, 271-320.

HECKMAN, J.J. AND E.J. VYTLACIL (1998): "Instrumental Variabl Methods for the Correlated Random Coefficients Model: Estimating the Average Returns to Schooling when the Return is Correlated with Schooling," *The Journal of Human Resources*, 33, pp. 974-1002.

HECKMAN, J.J. and E. VYTLACIL (2005) "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669-738.

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*.

HODERLEIN, S. AND E. MAMMEN (2007) "Identification of Marginal Effects in Nonseparable Models Without Monotonicity," *Econometrica*, Vol. 75, No. 5, pp. 1513-1518..

HOUTHAKKER, H.S. (1950) "Revealed Preference and the Utility Function," *Economica*, New Series, Vol. 17, No. 66, pp. 159-174.

KOOPMANS, T.C. (1950) *Statistical Inference in Dynamic Economic Models*, Cowles Commission for Research in Economics, Monograph No. 10. New York: John Wiley & Sons.

HSIAO, C. (1983), "Identification," in *Handbook of Econometrics*, Vol. 1, edited by Z. Griliches and M.D. Intriligator, North-Holland Publishing Company.

HURWICZ, L. (1950) "Generalization of the Concept of Identification," in *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10. New York: John Wiley.

HURWICZ, L. (1950) "Systems with Nonadditive Distrubances," in *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10. New York: John Wiley.

IMBENS, G.W. AND W.K. NEWHEY (2003) "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," mimeo, M.I.T.

KOOPMANS, T.C. (1949). "Identification problems in economic model construction". *Econometrica* 17 (2), 125-144.

KOOPMANS, T.C. and O. REIERSOL (1950) "The Identification of Structural Characteristics," *Annals of Mathematical Statistics*, 21, 165-181.

KOOPMANS, T.C., A. RUBIN, and R.B. LEIPNIK (1950) "Measuring the Equation System of Dynamic Economics," in *Statistical Inference in Dynamic Equilibrium Models*, Cowles Commission Monograph 10, New York: John Wiley.

KOTLARSKI, I. (1967) "On Characterizing the Gamma and Normal Distribution," *Pacific Journal of Mathematics*, Vol. 20, 69-76.

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

MALINVAUD, E. (1988) "Econometric Methodology at the Cowles Commission: Rise and Maturity [Address to the 50th Anniversary Cowles Celebration: 1983]" *Econometric Theory*, Vol. 4, No. 2, pp. 187-209.

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

MANSKI, C.F. (1989) "Anatomy of the Selection Problem," *Journal of Human Resources*, 24, 3, pp. 343-360.

MANSKI, C.F. (1995) *Identification Problems in the Social Sciences*, Cambridge, MA: Harvard University Press.

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

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

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

MATZKIN, R.L. (1999) "Nonparametric Estimation of Nonadditive Random Functions," mimeo, <http://www.econ.ucla.edu/faculty/?b=Matzkin>

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

MATZKIN, R.L. (2004) "Unobservable Instruments," mimeo, <http://www.econ.ucla.edu/faculty/?b=Matzkin>

MATZKIN, R.L. (2005a) "Identification of Consumers' Preferences when Individuals' Choices are Unobservable," *Economic Theory*, 26 (2), 423-443.

MATZKIN, R.L. (2005b) "Identification in Nonparametric Simultaneous Equations Models," mimeo, <http://www.econ.ucla.edu/faculty/?b=Matzkin>

MATZKIN, R.L. (2007a). "Heterogeneous choice". In Blundell, R., Newey, W.K., Persson, T. (eds.) *Advances in Economics and Econometrics; Theory and Applications, Ninth World Congress of the Econometric Society*, Vol. III. Cambridge University Press. New York.

MATZKIN, R.L. (2007). "Nonparametric Identification," Chapter 73 in *Handbook of Econometrics*, Vol. 6, Part B, edited by J.J. Heckman and E.E. Leamer, Elsevier.

MATZKIN, R.L. (2008), "Identification in Nonparametric Simultaneous Equations Models," *Econometrica*, Vol. 76, No. 5, pp. 945-978.

MATZKIN, R.L. (2010, 2013), "Estimation of Nonparametric Models with Simultaneity," mimeo, UCLA, <http://www.econ.ucla.edu/faculty/?b=Matzkin>

MATZKIN, R.L. (2012) "Identification in Nonparametric Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity," *Journal of Econometrics*, Vol. 166, No. 1, pp. 106-115.

McFADDEN, D. (1974) "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, ed. *Frontiers in Econometrics*, pp105-142, New York: Academic Press.

MUNDLAK, Y. (1963) "Estimation of Production Functions From a Combination of Cross Section and Time-Series Data," in *Measurement in Economics, Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, ed. by C. Christ. Stanford: Stanford University Press, PP. 138-166.

NG, S., PINSKE, J. (1995). "Nonparametric-two-step estimation of unknown regression functions when the regressors and the regression error are not independent." *Cahiers de recherche* 9551. Universite de Montreal, Departement de sciences economiques.

NEWHEY, W. and J. POWELL (1989) "Instrumental Variables Estimation of Nonparametric Models," mimeo.

NEWHEY, W. and J. POWELL (2003) "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*,.

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

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

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

ROEHRIG, C.S. (1988) "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56:2, 433-447.

ROTHENBERG, T.J. (1971) "Identification in Parametric Models," *Econometrica*, 39, 577-592.

SAMUELSON, P.A. (1938) "A Note on the Pure Theory of Consumer's Behaviour," *Economica*, New Series, Vol. 5, No. 17, pp. 61-71

STROTZ, R.H. and H.O.A. WOLD (1960) "Recursive vs. Nonrecursive Systems: An Attempt at Synthesis (Part 1 of a Triptych on Causal Systems)," *Econometrica*, 28, 417-463.

TAMER, E. (2010) "Partial Identification in Econometrics," *Annual Review of Economics*, No. 2, pp. 167-195.

TINBERGEN, J. (1930) "Bestimmung und Deutung von Angebotskurven: Ein Beispiel," *Zeitschrift für Nationalökonomie*, 70, 331-342.

WEGGE, L.L. (1965) "Identifiability Criteria for a System of Equations as a Whole," *The Australian Journal of Statistics*, 7, 67-77.