

Nonparametric Identification

Rosa L. Matzkin
Department of Economics
University of California - Los Angeles

Econ 312, Spring 2021

Introduction

- This chapter presents some of the recent results on the identification of nonparametric econometric models, concentrating on nonadditive models.
- It complements many current existent surveys that cover nonparametric identification, such as the books by Horowitz (1998), Pagan and Ullah (1999), and Yatchew (2003), articles in recent volumes of this Handbook by Härdle and Linton (1994), Matzkin (1994), Powell (1994), and van den Berg (2001), recent survey articles on semiparametric and nonparametric identification, such as Blundell and Powell (2003), Florens (2003), and Chesher (2007), and other chapters in this volume, such as the ones by X.Chen, Heckman and Vytlacil, and Ridder and Moffit.

- The objective of this chapter is to provide insight into some recent techniques that have been developed to identify nonparametric models, rather than on presenting a complete survey of the literature.
- As a consequence, many very important related works have been left out of the presentation and the references.

- When estimating an element in a model, it is necessary to determine first the identification of such an element.
- The study of identification in parametric econometric models dates back to the works by Workings (1925, 1927), Tinbergen (1930), Frisch (1934, 1938), Haavelmo (1943, 1944), Koopmans (1949), Hurwicz (1950), Koopmans and Reiersol (1950), Koopmans, Rubin and Leipnik (1950), Wald (1950), Fisher (1959, 1965, 1966), Wegge (1965), Rothenberg (1971), and Bowden (1973).
- (See Hausman (1983) and Hsiao (1983) in Volume 1 of this Handbook, for early review articles.)

- Lately, the analysis of identification in econometric models has been developing in several directions.
- One of these directions is the econometric analysis of systems of equations that require few or no parametric assumptions on the functions and distributions in the system.
- All the recent review articles mentioned above treat this topic.

- Imposing parametric specifications for functions and distributions had been the standard procedure in a world where large data sets were rarely available and computers could not easily handle estimation methods that require complicated computational algorithms.
- In such a world, estimating models with only a few parameters was part of the standard procedure.
- As computers processing power became faster and cheaper and the availability to deal with large data sets increased, it became possible to consider estimation of increasingly complicated functions, with increasing numbers of parameters.

- This, in turn, drove attention to the analysis of identification of functions and distributions that do not necessarily belong to parametric families.
- The emphasis was originally on estimation of probability densities and conditional expectations, but, later, more complicated models were considered.
- Rather than asking whether some parameters were identified, the question of interest became whether a function or distribution was identified within a general set of functions or distributions.
- Establishing such a nonparametric identification was recognized as an important first step in the econometric analysis of even parametric models.

- Establishing that a function or distribution is nonparametrically identified within a set of nonparametric functions or distributions implies its identification within any subset of the set of nonparametric functions.
- In particular, if the subset is defined as the set of functions that satisfy a parametric structure, such as being linear or quadratic, then identification within these subset is implied by identification within the larger set of nonparametric functions that include linear, quadratic, and possibly many other parametric specifications.
- If, on the other hand, one does not know whether the function is nonparametrically identified but one can establish its identification when a particular specification is imposed on the function, then it is not clear how robust any estimation results would be.

- When a function is nonparametrically identified, one can develop tests for different parametric structures, by comparing the results obtained from a nonparametric estimator for the function with those obtained from specific parametric estimators (Wooldridge (1992), Hong and White (1995), and Fan and Li (1996) are examples of such tests.)
- When a function is nonparametrically identified, one can allow the function to possess local behavior that would not be possible under some parametric specifications.
- (See, for example, the examples in Härdle (1991).)
- When a model or a function within a model is not identified nonparametrically, one can consider imposing sequentially stronger sets of restrictions in the model, up to the point where identification is achieved.

- This provides a method for analyzing the trade-off between imposing restrictions and achieving identification.
- (See, for example, Matzkin (1994) for such an analysis.)
- This chapter will present several of the developments in the nonparametric identification in economic models.

- Another area of active research, specially in recent years, was in the development of econometric models that were specified with properties closer to those of models studied in economic theory.
- The analysis of identification in the past, which concentrated on models that were linear in variables and parameters and additive in unobservable random terms, contrasted strongly with the standard practice in economic theory, where functions were only specified to possess some properties, such as continuity or monotonicity.
- On those times, economic theorists would work on models involving very general functions and distributions.
- Econometricians, on the other side, would work on models with well specified and typically quite restrictive functional forms and distributions.

- Even though the main goals of both groups were in many instances very similar, the solutions as well as the languages used in each of them were very different.
- The picture is drastically different nowadays.
- The development of nonparametric techniques for the estimation and testing of economic models has been shortening the distance between those roads to the point where now some econometric models are specified with no more restrictions than those that a theorist would impose.

- The advances that have decreased the distance between economic theory and econometrics have not concentrated only on the relaxation of parametric structures.
- Lately, there has also been an increasing effort to relax the way in which the unobservable random terms are treated.
- A practice that has been and still is commonly used when specifying an econometric model proceeds by first using economic theory to specify a relationship between a vector of observable explanatory variables and a vector of dependent variables, and then adding unobservable random variables to the relationships, as an after-thought.

- The seminal works by Heckman (1974), McFadden (1974), Heckman and Willis (1977), and Lancaster (1979) have shown that one can analyze econometric models where the unobservable random terms have important economic interpretations.
- They may represent, for example, heterogeneity parameters in utility functions, productivity shocks in production functions, or utility values for unobserved product attributes.
- When interpreting the unobservables in this way, it is rarely the case that they enter in additive ways into the models of interest.
- Several recent papers have considered the identification and estimation of nonparametric models with nonadditive random terms.
- Some of these will be reviewed in this chapter.

- Ideally, one would like to be able to identify all the unknown functions and distributions in a model without imposing more restrictions than those implied by the theory of the model.
- Restrictions derived from optimization, such as concavity and linear homogeneity, or equilibrium conditions, have been shown to be useful to identify functions in models that had been thought in the past to be identified only under very restrictive parametric assumptions.
- (See the survey chapter by Matzkin (1994) in Volume 4 of this Handbook for several such examples.)
- Nevertheless, in some cases, the identification of all functions and distributions in a model that imposes so few restrictions might not be possible.

- In such cases, one may consider several options.
- One may try to determine what can be identified without imposing any more restrictions on the model.
- One may impose some additional restrictions on some of the functions or distributions, to achieve identification.
- Or, one may consider enlarging the model, by augmenting the set of observable variables that can provide information about the functions or distributions of interest in the model.
- In this chapter we discuss some of the recent related techniques that have been developed.

- While restrictions implied by economic theory may, in some cases, aid in achieving identification, in some other cases, they may also hinder identification.
- This occurs when restrictions such as agent's optimization and equilibrium conditions generate interrelationships among observable variables, X , and unobservable variables, ε , that affect a common observable outcome variable, Y .
- In such cases, the joint distribution of (Y, X) does not provide enough information to recover the causal effect of X on Y , since changes in X do not leave the value of ε fixed.

- A typical example of this is when Y denotes quantity demanded for a product, X denotes the price of the product, and ε is an unobservable demand shifter.
- If the price that will make firms produce a certain quantity increases with quantity, this change in ε will generate an increment in the price X .
- Hence, the observable effect of a change in price in demanded quantity would not correspond to the effect of changing the value of price when the value ε stays constant.
- Another typical example arises when analyzing the effect of years of education on wages.

- An unobservable variable, such as ability, affects wages but also years of education.
- When an individual chooses years of education to maximize the discounted stream of future income, he takes ability into account because it influences the productivity of education.
- (See, for example, Card (2001).)
- As a result of this connection between ability and years of education, the distribution of ability, given years of education, changes with the years of education.
- In this chapter, we will review some of the methods that have been developed to identify causal effects in these situations.

- The outline of the chapter is as follows.
- In the next section, we describe several econometric models.
- In Section 3, we analyze, in general terms, identification in those models.
- In Section 4 we discuss some particular techniques that have been used to achieve identification.
- Section 5 concludes.

The econometric model

From the economic model to the econometric model

- The description of an economic model typically starts out by describing the economic agents involved, their objective functions, their information, and the interactions among the agents.
- When an econometrician tries to fit an economic model to the available data, he first needs to determine which of the variables in the model are observable and which are unobservable.
- Another important division of the variables in the model is between the variables that are determined outside of the model and those that are determined inside the model.
- The variables in the latter set are functions of the variables in the former set.
- In economic models, they are typically determined either by the choice of some agents or by the interaction among several agents.

- We will denote by X the vector of variables that are determined outside the model and are observable, and by ε the vector of variables that are determined outside the model and are unobservable.
- X and ε are also called the observable and unobservable explanatory, or exogenous, variables.
- We will denote the number of coordinates of X by K and the number of coordinates of ε by L .
- The vectors of observable and unobservable variables that are determined within the model will be denoted, respectively, by Y and Υ .
- These are observable and unobservable outcome variables.
- We will denote the number of coordinates in the vector of observable variables, Y , determined within the model, by G , and the number of coordinates in the vector of unobservable variables, Υ , determined within the model by G^Υ .

- Following the standard terminology, we will say that Y and Υ are vectors of, respectively, observable and unobservable endogenous variables.
- The description of an economic model contains, as well as a list of variables, a list of functions and distributions.

- Some of these functions and distributions are primitive, in the sense that they are determined outside the model.
- Some are derived within the model.
- Let \underline{h} denote the list of all primitive functions and let \underline{F} denote the list of all primitive distributions.
- We will describe the interrelation between the primitive functions and distributions and the observable and unobservable variables by a known vector function v and an equation

$$v(Y, \Upsilon, X, \varepsilon; \underline{h}, \underline{F}) = 0$$

- This equation can be used to derive the joint distribution of the vector of observable variables, (Y, X) , as a function of the primitives of the model, $(\underline{h}, \underline{F})$.

- To provide an example, consider a model of consumer demand for a consumption good and a composite good.
- Let I denote the income that the consumer can spend on these two goods.
- Let the price of the composite good be 1 and let p denote the price of the consumption good.
- Let y and z denote the quantities chosen by the consumer of, respectively, the consumption good and the composite good.
- Suppose that the economic model specifies that the individual has preferences over bundles (y, z) , and chooses the one that maximizes those preferences over the set of all bundles that cost no more than I .
- Suppose, further, that the consumer preferences can be represented by a strictly increasing, strictly concave, twice differentiable utility function, U , on (y, z) , and that such utility function is different for different individuals in a population.

- In particular, assume that the utility function depends on observable socioeconomic characteristics of the individual, such as age and marital status, denoted by w , and on unobservable tastes for (y, z) , denoted by ε .
- Then, for an individual with characteristics w and ε , and with observable income I , the observed choice (y, z) is defined as

$$(y, z) = \arg \max_{(\tilde{y}, \tilde{z})} \{U(\tilde{y}, \tilde{z}, w, \varepsilon) \mid p\tilde{y} + \tilde{z} \leq I\}$$

- Since the monotonicity of U with respect to (\tilde{y}, \tilde{z}) implies that all the available income will be used, this is equivalent to

$$y = \arg \max_{\tilde{y}} \{U(\tilde{y}, I - p\tilde{y}, w, \varepsilon)\}$$

$$z = I - py$$

- The differentiability, strict concavity, and strict monotonicity of U imply then that y satisfies

$$U_{\tilde{y}}(y, I - py, w, \varepsilon) - p U_{\tilde{z}}(y, I - py, w, \varepsilon) = 0$$

- In this model, the income, I , the vector of socioeconomic variables, w , and the price p are observable variables determined outside the system.
- The unobservable taste, ε , is also determined outside the system.
- The chosen quantity, y , of the commodity is observed and determined within the system.
- The utility function $U(\cdot, \cdot, \cdot, \cdot)$ is an unknown primitive function; and the distribution of (p, I, w, ε) is an unknown primitive distribution function.
- Given any particular utility function U , satisfying the differentiability, monotonicity and concavity restrictions imposed above, and given any distribution for (p, I, w, ε) , one can use the above equation to derive the joint distribution of the vector of observable variables, (Y, p, I, w) .

- This is derived from the equation

$$\begin{aligned}
 v(Y, X, \varepsilon) &= v(Y, p, I, w, \varepsilon) \\
 &= U_y(Y, I - pY, w, \varepsilon) - U_z(Y, I - pY, w, \varepsilon) p \\
 &= 0
 \end{aligned}$$

- Under our assumptions, the value of Y that satisfies this equation, for given values of (p, I, w, ε) , is unique.
- Let m denote the function that assigns the optimal value of Y to (p, I, w, ε) .
- Then, the demand function $m(p, I, w, \varepsilon)$ satisfies the first order conditions

$$\begin{aligned}
 &U_y(m(p, I, w, \varepsilon), I - pm(p, I, \varepsilon), w, \varepsilon) \\
 &\quad - U_z(m(p, I, w, \varepsilon), I - pm(p, I, \varepsilon), w, \varepsilon) p = 0
 \end{aligned}$$

- The demand model

$$Y = m(p, I, w, \varepsilon)$$

is the *reduced form* model.

- The reduced form model maps the observable and unobservable explanatory variables into the observable endogenous variables, without necessarily specifying behavioral and equilibrium conditions from which the mapping might have been derived.
- The reduced form model suffices to analyze many situations where this underlying structure does not change.
- For example, as will be discussed in more detail below, when m is strictly increasing in ε and ε is distributed independently of (p, I, w) , the reduced model above suffices to analyze the causal effect of (p, I, w) on Y .
- This is the effect on demand from changing the value of (p, I, w) , leaving the value of ε unchanged.

- The analysis of counterfactuals, on the other hand, would typically require knowledge of the primitive function U .
- Suppose, for example, that we were interested in predicting the behavior of a consumer that possesses preferences as in the model above, when the price of the consumption good depends on the quantity chosen, instead of being a fixed value, p , as considered above.
- Denote the price function as $s(y)$.
- To predict the choice of the consumer with utility function $U(\tilde{y}, \tilde{z}, w, \varepsilon)$ when his set of affordable consumption bundles is

$$\{(\tilde{y}, \tilde{z}) \mid s(\tilde{y}) \tilde{y} + \tilde{z} = I\}$$

we would need to know the function $U(\tilde{y}, \tilde{z}, w, \varepsilon)$ to calculate the new optimal values

$$(y, z) = \arg \max_{(\tilde{y}, \tilde{z})} \{U(\tilde{y}, \tilde{z}, w, \varepsilon) \mid s(\tilde{y}) \tilde{y} + \tilde{z} = I\}$$

- This would require analyzing the *structural model* of utility maximization described earlier.
- The structural model uses behavioral and/or equilibrium conditions, to define a mapping between the primitive functions and distributions, on one side, and the distribution of the observable variables, on the other.
- Path diagrams (Pearl (2000)) are often very useful to clarify the role of each variable and the ordering of the variables in terms of cause and effect.
- Support conditions, which may allow one to identify only the local behavior of some functions should also be taken into consideration.

Dependence between ε and X

- In many cases, a model is not completely specified.
- Some of the unobservable explanatory variables in the model are themselves functions of observable variables, in a way that is not described within the model.
- Consider, for example, the utility maximization model described in the previous subsection.
- In that model, the income of the consumer, I , was assumed to be determined outside of the model.
- The unobservable ε was assumed to denote taste for consumption.
- In many cases, one could think of income as being partially determined by ε .
- Individuals with a larger taste for consumption will typically make lifetime decisions, such as the choice of profession, that would generate higher incomes.

- In particular, if we let \tilde{r} denote a function and let δ denote additional variables, which are determined outside the system and which affect income I , we could specify that $I = \tilde{r}(\varepsilon, \delta)$.
- If this latter relationship were added to the specification of the model, then, in the augmented model, the variables determined within the system would be (Y, Z, I) , and those determined outside the system would be (p, ε, δ) .
- Suppose that we wanted to infer the causal effect of income I on demand Y .
- This is the effect on Y of changing I , when the value of (p, w, ε) stays fixed.
- If I is a function of ε , the total effect will be different from this partial effect.
- A similar example occurs when variables are determined jointly.
- Haavelmo (1943, 1944) argued that in these cases a joint probability distribution is needed to analyze the data.

2.2. Definition of an econometric model

- Following up on the model described in the beginning of Section 2, we define an *econometric model* by a specification of variables that are observed and variables that are unobserved, variables that are determined within the model and variables that are determined outside of the model, functional relationships among all the variables, and restrictions on the functions and distributions.
- We will denote by S the set of all vectors of functions and distributions that satisfy the restrictions imposed by the model.
- We assume that for any element $\zeta \in S$, we can derive the distribution, $F_{Y,X}(\cdot; \zeta)$, of the observable vector of variables that is generated by S .
- The observable distribution, $F_{Y,X}$, corresponds to the true value ζ^* of ζ .

- For example, in the consumer demand model described above, ε and (p, I, w) are, respectively, the vectors of unobservable and observable explanatory variables and Y is the vector of observable endogenous variables.
- The elements of S are pairs $\zeta = (U, F_{\varepsilon, p, I, w})$, such that for all (w, ε) , $U(\cdot, \cdot, w, \varepsilon) : R^2 \rightarrow R$ is strictly increasing, strictly concave, and twice differentiable, and $F_{\varepsilon, p, I, w}$ is a distribution function.
- Given $\zeta = (U, F_{\varepsilon, p, I, w})$ and $X = (p, I, w)$, the distribution of Y given X is calculated by the distribution of ε given (p, I, w) and the function U , using the first order conditions.
- Note that since X is observable, the marginal distribution of X , F_X , can be assumed to be known.
- Hence, one of the restrictions that $F_{\varepsilon, p, I, w}$ would be required to satisfy is that the marginal distribution of (p, I, w) coincides with $F_{p, I, w}$.

Examples

- We next describe several models, whose identification will be discussed in Sections 3 and 4.
- We denote random variables with capital letters and their realizations with lower case letters.

2.2.1.1 Additive Models

- In additive models, the unobservable variables that are determined outside the model affect the values of the variables that are determined within the model in an additive way.
- A standard example of such a model is where Y denotes an observable dependent variable, X denote a vector of observable explanatory variables, ε denotes an unobservable explanatory variable, and the functional relationship between these variables is given by

$$Y = X\beta + \varepsilon$$

for some β .

- Allowing X to influence Y in a nonlinear, possibly unknown way, while leaving the influence of ε additive, will also give rise to an additive model.

- In this latter case

$$Y = g(X) + \varepsilon$$

for some function g .

- Typical restrictions that are imposed on such a model are that g is continuous and that the distribution of ε given X has support R .
- Typically, one would like to add the restriction that the distribution of (X, ε) is such that for all x in some set, the conditional expectation of ε given $X = x$ is 0.

- In such a case $g(x)$ denotes the conditional expectation of Y given $X = x$, which is an object of interest when forecasting the value of Y conditional on $X = x$, under a squared-error loss function.
- In other situations, one may want to add the restriction that the conditional median, or other quantile of ε , given $X = x$ is 0.
- Many methods exist to estimate conditional means and conditional quantiles nonparametrically. Prakasa Rao (1983), Härdle and Linton (1994), Pagan and Ullah (1999), Matzkin (1994), Koenker (2005), and X. Chen (2007), among others, survey parts of this literature.

2.2.1.2 Nonadditive Models

- When the unobservable random terms in an economic model have important interpretations such as being variables representing tastes of consumers, or productivity shocks in production functions, it is rarely the case that these unobservable random terms influence the dependent variables in the model in an additive way.
- Nonadditive models allow the unobservable variables that are determined outside the model to affect the values of the variables that are determined within the model in nonadditive ways.

- For a simple example, let Y denote an observable dependent variable, X denote a vector of observable explanatory variables, and ε denote an unobservable explanatory variable.
- We can specify the functional relationship between these variables as

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

for some function $m : R^K \times R \rightarrow R$.

- We may impose the restrictions that the function m is strictly increasing in ε , for all values of X , and that the distribution, $F_{\varepsilon, X}$, of (X, ε) is strictly increasing over R^{K+1} .
- We may add the restriction that m is differentiable, or that X and ε are distributed independently of each other.
- When the latter restriction is imposed, we will call such model an *Independent Nonadditive Model*.

- An example of such a model could be when X denotes hours of work of an individual, ε denotes the ability of the individual to perform some task, and Y is output of the individual.
- Conditional on working the same quantity x of hours of work, output is higher when ability is higher.

- Nonparametric models of this type were studied in Roehrig (1988), Olley and Pakes (1996), Brown and Matzkin (1998), Matzkin (1999, 2003), Altonji and Ichimura (2000), Altonji and Matzkin (2001), and Imbens and Newey (2003), among others.
- When the distribution of ε is specified to be $U(0, 1)$ and m is strictly increasing in ε , the function m can be interpreted as a nonparametric conditional quantile function.
- See Chaudhuri (1991) and Chaudhuri, Doksum, and Samarov (1997), for nonparametric estimation, as well as the references in Koenker (2005).

- The additive model described in Section 2.2.1.1 can be interpreted as a different representation of the nonadditive model.
- One can always express the model: $Y = m(X, \varepsilon)$ as $Y = g(X) + \eta$, where for each x , $g(x) = E(Y|X = x)$.
- In such case, the value of the additive unobservable η has, by construction, conditional expectation equal 0, given $X = x$.
- The distribution of η given $X = x$ can be derived from the function m and the distribution of ε given $X = x$, since by its definition, $\eta = Y - E(Y|X = x) = m(X, \varepsilon) - g(x)$.

2.2.1.3. Triangular Nonadditive Model

- When m and ε are multivalued, a particular nonadditive model is the *Triangular Nonadditive Model*.
- In this model, there are G endogenous (outcome) variables, Y_1, \dots, Y_G , and G unobservable variables, $\varepsilon_1, \dots, \varepsilon_G$.

- Given a vector of explanatory variables, $X \in R^K$, the value of each Y_g is determined recursively from X , Y_1, \dots, Y_{g-1} , and ε_g :

$$Y_1 = m_1(X, \varepsilon_1)$$

$$Y_2 = m_2(X, Y_1, \varepsilon_2)$$

$$Y_3 = m_3(X, Y_1, Y_2, \varepsilon_3)$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$Y_G = m_G(X, Y_1, Y_2, \dots, Y_{G-1}, \varepsilon_G)$$

- This is a nonparametric nonadditive version of the triangular system in linear simultaneous equations (see Hausman (1983)), where for some lower triangular, $G \times G$ matrix A and some $G \times K$ matrix B ,

$$\varepsilon = AY + BX$$

where ε is the $G \times 1$ vector $(\varepsilon_1, \dots, \varepsilon_G)'$, Y is the $G \times 1$ vector $(Y_1, \dots, Y_G)'$, and X is the $K \times 1$ vector $(X_1, \dots, X_K)'$.

- Nonparametric identification in the nonparametric, nonadditive model has been studied recently by Chesher (2003) and Imbens and Newey (2003), among others.
- The later considers also nonparametric estimation.
- (Ma and Koenker (2006) compare the approaches of those two papers. See also Matzkin (2004)).
- A typical example (see Imbens and Newey (2003) and Chesher (2003)) is the model where Y_2 denotes lifetime discounted income, Y_1 denotes years of education, X is a variable denoting the cost of education, ε_1 is (unobserved) ability, and ε_2 is another unobservable variable that affects income.
- In this example, X is an argument of the function m_1 but not of the function m_2 .
- Many panel data models, where the unobservables incorporate fixed effects, fall into this structure.

- By recursively substituting the endogenous variables, in the above system of the equations, one can obtain the system of reduced form equations, where each endogenous variable is solely determined by observable and unobservable exogenous variables.

- This system has the form

$$Y_1 = h_1(X, \varepsilon_1)$$

$$Y_2 = h_2(X, \varepsilon_1, \varepsilon_2)$$

$$Y_3 = h_3(X, \varepsilon_1, \varepsilon_2, \varepsilon_3)$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$Y_G = h_G(X, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_G)$$

where $h_1(X, \varepsilon_1) = m_1(X, \varepsilon_1)$, $h_2(X, \varepsilon_1, \varepsilon_2) = m_2(X, Y_1, \varepsilon_1, \varepsilon_2) = m_2(X, h_1(X, \varepsilon_1), \varepsilon_1, \varepsilon_2)$, and so on.

- As can be seen from above, the reduced form of this model, which represents the G dimensional vector of outcomes Y_1, \dots, Y_G as G functions of the vector of observable explanatory variables, X , and the vector of G unobservable variables $\varepsilon_1, \dots, \varepsilon_G$, is triangular in $(\varepsilon_1, \dots, \varepsilon_G)$, in the sense that for each g , Y_g does not depend on $\varepsilon_{g+1}, \dots, \varepsilon_G$.

2.2.1.4. Nonadditive Index Models

- In many situations in economics, we might be interested in analyzing the effect that some vector of variables X has on a variable, Y , when the model establishing such a relationship between X and Y is either very complicated or only vaguely known.
- If we could determine that the effect of X on Y is weakly separable from the other variables, then we might be able to identify features of the aggregator, or "index" function, $h(X)$, even though we might not be able to infer all the functions and distributions in the model.

- A simple example of a nonadditive index model is where Y denotes an observable dependent variable, X denotes a vector of observable explanatory variables, and ε denotes an unobservable explanatory variable.
- The functional relationship between these variables is specified as

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

where $m : R^2 \rightarrow R$ and $h : R^K \rightarrow R$.

- We may impose the restrictions that m is increasing in each coordinate and h is continuous.

- Stoker (1986), Han (1987), Powell, Stock and Stoker (1989), Ichimura (1993), Horowitz (1996), Horowitz and Hardle (1996), Abrevaya (2000), and Das (2001) have considered semiparametric estimation of single index linear models, where the function h is specified as a linear-in-parameters function.
- Ichimura and Lee (1991) considered identification and estimation of semiparametric, multiple linear index models.
- Matzkin (1991b) considered estimation of a nonparametric h .
- Matzkin and Newey (1993), Horowitz (2001), and Lewbel and Linton (2007) considered estimation of h and the distribution of ε nonparametrically.
- Heckman and Vytlacil (1999, 2000) and Vytlacil and Yildiz (2004), among others, consider identification of average effects.
- Chesher (2005) considers local identification when X is endogenous and ε is vector valued.

- If we impose the restriction that X and ε are independently distributed, we will call it the *Independent Nonadditive Index Model*.
- Consider, for example, a duration model, with a proportional hazard function, $\lambda(t, x, \nu)$, given by

$$\lambda(t, x, \nu) = s(t) e^{h(x)+\nu}$$

where x denotes the value of observable characteristics, X , ν denotes the value of an unobservable characteristic, and t denotes the time, Y , at which the hazard is evaluated.

- Suppose that r is an unknown positive function over R_+ , h is an unknown function over the support of X , and ν is distributed independently of X .

- Such a model could describe a situation where Y denotes the length of time that it takes an individual with observable characteristics, X , and unobservable characteristic, ν , to find employment.
- When the probability-density of finding employment at time t conditional on not having found employment yet is given by the above specification for the hazard function, the model that describes the relation between Y and X is

$$Y = m(h(X), \eta + \nu)$$

where η possesses an extreme value distribution, independent of (X, ν) .

- Moreover, m is strictly decreasing in $\eta + \nu$.

- Semiparametric and nonparametric identification of duration models, as well as corresponding estimation methods, were studied by Elbers and Ridder (1982), Heckman (1991), Heckman and Singer (1984a, 1984b), Barros and Honore (1988), Honore (1990), Ridder (1990), Horowitz (1999), van den Berg (2001), and Abbring and van der Berg (2003).
- (See also the chapters on this topic in Lancaster (1990).)

2.2.1.5. Nonadditive Simultaneous Equations Models

- In many economic models the values of the dependent variables are determined simultaneously.
- A standard example is the model of demand and supply.
- Let m^d denote an aggregate demand function, which determines the aggregate quantity demanded of a product, Q^d , as a function of the price of the product, p , the income level of the consumers, I , and an unobservable variable ε^d .
- Let m^s denote the aggregate supply function, which determines the aggregate supplied output, Q^s , as a function of the price of the product, P , input prices, W , and an unobservable variable, ε^s .

- In equilibrium, $Q^d = Q^s$.
- The model can then be described as

$$Q^d = m^d(P, I, \varepsilon^d)$$

$$Q^s = m^s(P, W, \varepsilon^s)$$

$$Q^d = Q^s$$

where the last equation denotes the equilibrium conditions that aggregate demand equals aggregate supply.

- In this model, the equilibrium quantity, $Q = Q^d = Q^s$, and the equilibrium price are determined simultaneously.
- In most multidimensional optimization problems, such as those faced by a consumer maximizing a utility function or by a multiproduct firm maximizing profits, the optimal choices are also determined simultaneously.

- The analysis of simultaneous equations models is typically more complicated than that of many other models because the unobservables that affect any one of the endogenous variables affect, through the simultaneity, also the other endogenous variables.
- This was made clear for linear models by Haavelmo (1943), who showed that Least Squares was not the correct method to estimate models with endogenous variables.
- Suppose, for example, that in the demand and supply example described above, m^d is strictly increasing in ε^d and m^s is strictly decreasing in ε^s .

- Then, the system can be expressed as

$$\begin{aligned}\varepsilon^d &= r^d(Q, P, I) \\ \varepsilon^s &= r^s(Q, P, W)\end{aligned}$$

where r^d is the inverse function of m^d with respect to ε^d and r^s is the inverse function of m^s with respect to ε^s .

- Assuming that, for any value of the vector of exogenous variables, $(I, W, \varepsilon^d, \varepsilon^s)$, this system of structural equations possesses a unique solution for (P, Q) , one can derive the reduced form system of the model, which can be expressed as

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

- When the structural equations in the simultaneous equations model above are linear in the variables, as in the standard linear models for simultaneous equations, the reduced form equations turn out to be linear in the unobservables.
- In such case, to each reduced form equation there corresponds a unique unobservable random term, which enters the equation in an additive way.
- The value of each such unobservable is a function of ε^d , ε^s and of the coefficients that appear in r^d and r^s .
- Identification in linear simultaneous equations can be analyzed using the results in Koopmans (1949), Koopmans, Rubin, and Leipnik (1950), and Fisher (1966), among others.
- (See Hausman (1983) and Hsiao (1983) for surveys of that literature.)

- We will consider below the nonadditive simultaneous equations model described by

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

where $Y \in R^G$ denote a vector of observable dependent variables, $X \in R^K$ denote a vector of observable explanatory variables, and $\varepsilon \in R^L$ denote a vector of unobservable explanatory variables.

- The function $r : R^G \times R^K \rightarrow R^L$ specifies the relationship between these vectors.
- In our analysis of this model, we will impose the restriction that r is differentiable and is such that for all values of (X, ε) there is a unique Y satisfying the above equation.
- We will also impose the restriction that X and ε are independently distributed with support $R^K \times R^G$ and that r is such that for each x , the density of Y given $X = x$ has support R^G .

- The identification of nonparametric simultaneous equations satisfying these properties was first analyzed by Roehrig (1988), following a technique developed by B.
- Brown (1983) for parametric, nonlinear in variables, simultaneous equations models.
- Recently, Benkard and Berry (2004) showed that Roehrig's conditions may not guarantee identification.
- Matzkin (2005) proposed a different set of conditions.
- Manski (1983) proposed a Closest Empirical Distribution method for estimation of a semiparametric version of these models, which did not require a parametric specification for the density of ε .

- Brown and Matzkin (1998) developed a nonparametric Closest Empirical Distribution method, which did not require either the distribution of ε or the function r to be parametric.
- A semi-nonparametric maximum likelihood method, such as that developed in Gallant and Nychka (1987), or a semiparametric maximum likelihood method, as in Ai (1997) could also be used to estimate identified models.

- When a structural function is additive in the unobservable random term, estimation can proceed using the nonparametric instrumental variable methods of Newey and Powell (1989, 2003), Ai and Chen (2003), Darolles, Florens, and Renault (2002), and Hall and Horowitz (2005).
- When it is nonadditive, the methods of Chernozhukov and Hansen (2005), or Chernozhukov, Imbens, and Newey (2007) could be used.

2.2.1.6. Discrete Choice Models

- Discrete Choice Models are models typically used to describe the situation where an individual has a finite number, $1, \dots, G$, of alternatives to choose from.
- The individual has preferences defined over those alternatives and chooses one that maximizes those preferences.
- It is assumed that the preference of the individual for each alternative can be represented by a function, V_g , which depends on observable and unobserved characteristics of the individual and of the alternative.
- Let S denote a vector of observable socioeconomic characteristics of a typical individual.
- Let Z_g denote a vector of observable characteristics of alternative g .

- Let ε denote a vector of unobservable variables.
- It is typically assumed that $\varepsilon \in R^J$ where $J \geq G$.
- For each g , let $Y_g^* = V_g(S, Z_g, \varepsilon)$, and let $Y_g = 1$ if the individual chooses alternative g and $Y_g = 0$ otherwise.
- Assume that the functions V_1, \dots, V_G and the distribution of ε are such that there is zero probability that for some $g \neq k$, $V_g(S, Z_g, \varepsilon) = V_k(S, Z_k, \varepsilon)$.

- In this model, the vector of unobserved endogenous variables is $Y^* = (Y_1^*, \dots, Y_G^*)$, and the vector of observable endogenous variables is $Y = (Y_1, \dots, Y_G)$ where, for each g ,

$$Y_g = \left\{ \begin{array}{ll} 1 & \text{if } V_g(S, Z_g, \varepsilon) > V_k(S, Z_k, \varepsilon) \quad \text{for all } k \neq g \\ 0 & \text{otherwise} \end{array} \right\}$$

- The vector of observable explanatory variables is $X = (S, Z_1, \dots, Z_G)$.

- The conditional probability of Y given X is given by

$$\Pr(Y_g = 1|X) = \Pr(\{\varepsilon|V_g(S, Z_g, \varepsilon) > V_k(S, Z_k, \varepsilon) \\ \text{for all } k \equiv j\})$$

- Discrete Choice Models were originally developed by McFadden (1974) under the linear additive specification that for all g

$$V_g(S, Z_g, \varepsilon) = \alpha_g + \gamma_g S + \beta_g Z_g + \varepsilon_g$$

and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_G)$.

- Initially, McFadden (1974) specified a parametric distribution for ε .
- Subsequent work by Manski (1975, 1985), Cosslett (1983), Powell, Stock and Stoker (1989), Horowitz (1992), Ichimura (1993), and Klein and Spady (1993), among others, developed methods that did not require a parametric specification for ε .
- Matzkin (1991a) considered identification when the distribution of $\varepsilon = (\varepsilon_1, \dots, \varepsilon_G)$ is specified parametrically and for each g

$$V_g(S, Z_g, \varepsilon) = v_g(S, Z_g) + \varepsilon_g$$

for some unknown functions v_g .

- Matzkin (1992, 1993) extended these results to the case where both the distribution of $(\varepsilon_1, \dots, \varepsilon_G)$ and the functions v_1, \dots, v_G are nonparametric.

Identification

Definition of identification

- Following the description of an econometric model in Section 2, we denote the set of all vectors of functions and distributions that satisfy the restrictions imposed by a model by S .
- We denote any element in S by ζ , and we denote the element of S corresponding to the vector of true functions and distributions by ζ^* .
- For any element ζ in S , we will denote by $F_{Y,X}(\cdot, \cdot; \zeta)$ the distribution of the observable variables generated by ζ .
- The distribution of the observable variables generated by ζ^* will be denoted by $F_{Y,X}(\cdot, \cdot; \zeta^*)$ or simply by $F_{Y,X}$.

- The analysis of identification deals with the mapping between the distribution of the observable variables and the underlying elements in the model.
- Given a model, with an associated vector of functions and distributions, ζ^* , and a set S of vectors of functions and distributions satisfying the same restrictions that ζ^* is assumed to satisfy, we can ask what elements of ζ^* are uniquely determined from $F_{Y,X}$.

- More generally, we may ask what features of ζ^* can be uniquely recovered from $F_{Y,X}$.
- By a *feature* of ζ , we mean any function $\Psi : S \rightarrow \Omega$.
- This could be an element of ζ , or a property such as, for example, the sign of the derivative of a particular function in ζ .
- We will let $\psi^* = \Psi(\zeta^*)$; ψ^* then denotes the true value of the feature of ζ^* .
- Elements in the range, $\Psi(S)$, of Ψ will be denoted by ψ .

- Given $\psi \in \Psi(S)$, we define $\Gamma_{Y,X}(\psi, S)$ to be the set of all probability distributions of (Y, X) that are consistent with ψ and S .
- Formally,

$$\Gamma_{Y,X}(\psi, S) = \{F_{Y,X}(\cdot, \cdot; \zeta) \mid \zeta \in S \text{ and } \Psi(\zeta) = \psi\}$$

- In other words, $\Gamma_{Y,X}(\psi, S)$ is the set of all distributions of (Y, X) that are generated by some vector of functions and distributions in S and whose value of the element that we want to infer is ψ .

- In the model of consumer demand, ψ^* may denote, for example, the utility function U^* , the expected demand of a socioeconomic group at a particular budget $E[m^*(p, I, w, \varepsilon) | p, I, w]$, or the expected infinitesimal effect in the demand of a change in price, $E[\partial m^*(p, I, w, \varepsilon) / \partial p | p, I, w]$.

- A key concept when analyzing identification is the one of observational equivalence.
- Two values $\psi, \psi' \in \Omega$ are observationally equivalent if there exist at least two vectors, $\zeta, \zeta' \in \mathcal{S}$ with $\Psi(\zeta) = \psi$, $\Psi(\zeta') = \psi'$, and $F_{Y,X}(\cdot, \cdot; \zeta) = F_{Y,X}(\cdot, \cdot; \zeta')$:

- **Definition 3.1:** $\psi, \psi' \in \Omega$ are *observationally equivalent* in the model S if

$$[\Gamma_{Y,X}(\psi, S) \cap \Gamma_{Y,X}(\psi', S)] \neq \emptyset$$

- The feature ψ^* is identified if there is no $\psi \in \Omega$ such that $\psi \neq \psi^*$ and ψ is observationally equivalent to ψ^* :

- **Definition 3.2:** $\psi^* \in \Omega$ is identified in model S if for any $\psi \in \Omega$ such that $\psi \neq \psi^*$

$$[\Gamma_{Y,X}(\psi, S) \cap \Gamma_{Y,X}(\psi^*, S)] = \emptyset$$

- The following characterization is often used to prove identification when it is easy to show that ψ^* can be recovered uniquely from any distribution in $\Gamma_{Y,X}(\psi^*, S)$ in particular models:

- **Definition 3.3:** $\psi^* \in \Omega$ is identified in model S if for any $\psi \in \Omega$

$$([\Gamma_{Y,X}(\psi, S) \cap \Gamma_{Y,X}(\psi^*, S)] \neq \emptyset) \Rightarrow [\psi = \psi^*]$$

Identification in Additive Models

- Consider the model

$$Y = g^*(X) + \varepsilon$$

where Y denotes an observable dependent variable, $X \in R^K$ denotes a vector of observable explanatory variables, ε denotes an unobservable explanatory variable, and $g^* : R^K \rightarrow R$ is an unknown, continuous function.

- Suppose that we were interested in the value $g^*(\bar{x})$ of the function g^* at a particular value \bar{x} of X .
- For any distribution $\tilde{F}_{\varepsilon, X}$ of (ε, X) , let $E[\varepsilon | X = x; \tilde{F}_{\varepsilon, X}]$ denote the expectation of ε conditional on $X = x$, calculated using $\tilde{F}_{\varepsilon, X}$, and let \tilde{f}_X denote the probability density of the marginal distribution \tilde{F}_X .

- Let $S = \{(\tilde{g}, \tilde{F}_{\varepsilon, X}) \mid \tilde{g} : R^K \rightarrow R \text{ is continuous and } \tilde{F}_{\varepsilon, X} \text{ is a distribution on } R^{K+1} \text{ such that (i) } \tilde{f}_X(\bar{x}) > 0 \text{ and } \tilde{f}_X \text{ has an extension that is continuous at } \bar{x}, \text{ (ii) } E[\varepsilon \mid X = \bar{x}; \tilde{F}_{\varepsilon, X}] = 0 \text{ and } E[\varepsilon \mid X = x; \tilde{F}_{\varepsilon, X}] \text{ has an extension that is continuous in } x \text{ at } \bar{x}\}$.
- Let Ω denote the set of all possible values that $\psi^* = g^*(\bar{x})$ can attain.
- Then,

$$(3.a) \quad \psi^* = g^*(\bar{x}) \text{ is identified}$$

- **Proof of (3.a):** Let $E \left[Y|X = x; \tilde{g}, \tilde{F}_{\varepsilon,X} \right]$ denote the conditional expectation of Y given $X = x$, for the distribution generated by $(\tilde{g}, \tilde{F}_{\varepsilon,X})$.
- Suppose that $(g^*, F'_{\varepsilon,X}), (\tilde{g}, \tilde{F}_{\varepsilon,X}) \in S$ and $\tilde{g}(\bar{x}) \neq g^*(\bar{x})$.
- Then, since

$$E \left[Y|X = \bar{x}; \tilde{g}, \tilde{F}_{\varepsilon,X} \right] = \tilde{g}(\bar{x}) + E \left[\varepsilon|X = \bar{x}; \tilde{F}_{\varepsilon,X} \right] = \tilde{g}(\bar{x})$$

$$E \left[Y|X = \bar{x}; g^*, F'_{\varepsilon,X} \right] = g^*(\bar{x}) + E \left[\varepsilon|X = \bar{x}; F'_{\varepsilon,X} \right] = g^*(\bar{x})$$

and both functions are continuous at \bar{x} , it follows by the properties of $F'_{\varepsilon,X}$ and $\tilde{F}_{\varepsilon,X}$ that

$$F_{Y,X}(\cdot; g^*, F'_{\varepsilon,X}) \neq F_{Y,X}(\cdot; \tilde{g}, \tilde{F}_{\varepsilon,X})$$

- Hence, ψ^* is identified. ■

- When g^* is identified, we can also identify $F_{\varepsilon, X}^*$.
- Assume for simplicity that the marginal distribution F_X has an everywhere positive density.
- Let $S = \{(\tilde{g}, \tilde{F}_{\varepsilon, X}) \mid \tilde{g} : R^K \rightarrow R \text{ is continuous and } \tilde{F}_{\varepsilon, X} \text{ is a distribution that has support } R^{K+1} \text{ and is such that } E[\varepsilon \mid X = x; \tilde{F}_{\varepsilon, X}] \text{ is continuous in } x \text{ and it equals 0 at all values of } x\}$.
- Let Ω denote the set of all possible pairs of functions $\psi = (g, F_{\varepsilon, X})$.
- Then,

$$(3.b) \quad \psi^* = (g^*, F_{\varepsilon, X}^*) \text{ is identified}$$

- **Proof of (3.b):** Using the same arguments as in the proof of (3.a), we can show that, for any x , $g^*(x)$ is identified.
- To show that $F_{\varepsilon, X}^*$ is identified, note that

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

- Since the marginal density, f_X^* , of X is identified, it follows that $F_{\varepsilon, X}^*(x, e)$ is identified. ■

- The linear model is, of course, the most well known case of an additive model.
- In this case, for all x ,

$$g^*(x) = \alpha^* + \beta^* x$$

for some $\alpha^* \in R$, $\beta^* \in R^K$.

- To identify $\psi^* = (\alpha^*, \beta^*)$ within the set of all vectors $(\alpha, \beta) \in R^{1+K}$, one needs a rank condition in addition to the location normalization.

- Suppose that for $K + 1$ vectors $x^{(1)}, \dots, x^{(K+1)}$, $g^*(x^{(k)})$ is identified and the rank of the $(K + 1) \times (K + 1)$ matrix whose k -th row is $(1, x^{(k)})$ is $K + 1$.
- Then, the system of $K + 1$ linear equations

$$\alpha^* + \beta^* x^{(k)} = g^*(x^{(k)}) \quad k = 1, \dots, K + 1$$

has a unique solution.

- Hence, (α^*, β^*) is identified.

3.3. Identification in Nonadditive Models

- Since the nonadditive model is more general than the additive model, it would not be surprising to find out that stronger conditions are necessary for the identification of the function m^* and distribution $F_{\varepsilon, X}^*$ in the model where Y is an observable dependent variable, X is a vector of observable explanatory variables, ε is an unobservable random term explanatory variable, and

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

- In fact, Matzkin (2003, Lemma 1) establishes that even when m^* is assumed to be strictly increasing in ε and ε is distributed independently of X , one can not identify m^* .
- Assume that F_X is known. Let Ξ denote the support of X , which will be assumed to be R^K .
- We will assume that F_ε^* has support R and that ε is distributed independently of X .
- Hence, we can characterize the model by pairs (m, F_ε) .

- **Theorem 3.1** (Matzkin (2003)): *Let*
 $S = \{(\tilde{m}, \tilde{F}_\varepsilon) \mid \tilde{m} : \Xi \times R \rightarrow R \text{ is continuous on } \Xi \times R \text{ and strictly increasing in its last coordinate and } \tilde{F}_\varepsilon \text{ is continuous and strictly increasing on } R\}$.
- *Let* $\Psi : S \rightarrow \Omega$ *denote the first coordinate of* $\zeta = (m, F_\varepsilon) \in S$.
- *Then,* $m, \tilde{m} \in \Omega$ *are observationally equivalent iff for some continuous and strictly increasing function* $s : R \rightarrow R$ *and all* $x \in \Xi, \varepsilon \in R$

$$\tilde{m}(x, s(\varepsilon)) = m(x, \varepsilon)$$

- **Proof:** Suppose $m, \tilde{m} \in \Omega$ are observationally equivalent.
- Then, there exist continuous and strictly increasing $F_\varepsilon, \tilde{F}_\varepsilon$ such that for all $x \in \Xi, y \in R$

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

- Let $r(x, \cdot)$ and $\tilde{r}(x, \cdot)$ denote, respectively, the inverses of $m(x, \cdot)$ and $\tilde{m}(x, \cdot)$.
- Since for all y, x

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

and

$$F_{Y|X=x}(y; (\tilde{m}, \tilde{F}_\varepsilon)) = \Pr(Y \leq y | X = x; (\tilde{m}, \tilde{F}_\varepsilon)) = \tilde{F}_\varepsilon(\tilde{r}(y, x))$$

it follows that for all y, x

$$F_\varepsilon(r(y, x)) = \tilde{F}_\varepsilon(\tilde{r}(y, x))$$

- Since F_{ε} , \tilde{F}_{ε} are strictly increasing and continuous, the function $s(t) = \tilde{F}_{\varepsilon}^{-1}(F_{\varepsilon}(t))$ is strictly increasing and continuous and $\tilde{r}(y, x) = s(r(y, x))$.
- Let $y = m(x, \varepsilon)$.
- Since \tilde{r} is the inverse of \tilde{m}

$$y = \tilde{m}(x, \tilde{r}(y, x)) = \tilde{m}(x, s(r(y, x))) = \tilde{m}(x, s(\varepsilon))$$

- Hence,

$$m(x, \varepsilon) = \tilde{m}(x, s(\varepsilon))$$

- Conversely, suppose that m and \tilde{m} are such that for a strictly increasing and continuous function s , all x and ε

$$m(x, \varepsilon) = \tilde{m}(x, s(\varepsilon))$$

- Let F_ε denote any continuous and strictly increasing distribution on R .
- Let $\tilde{\varepsilon} = s(\varepsilon)$ and let \tilde{F}_ε denote the distribution of $\tilde{\varepsilon}$, which is derived from s and F_ε .
- Let r and \tilde{r} denote respectively the inverse functions of m with respect to ε and of \tilde{m} with respect to $\tilde{\varepsilon}$.
- Then, for all y, x

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

and

$$F_{Y|X=x}(y; (\tilde{m}, \tilde{F}_\varepsilon)) = \Pr(Y \leq y | X = x; (\tilde{m}, \tilde{F}_\varepsilon)) = \tilde{F}_\varepsilon(\tilde{r}(y, x))$$

- Hence, m and \tilde{m} are observationally equivalent. ■

- An implication of the above result is that to identify m^* , one must restrict m^* to belong to a set of functions such that for any two different continuous functions in the set, their corresponding inverse functions are not continuous, strictly increasing transformations of each other.
- Suppose, for example, that we impose the normalization that for some \bar{x} for which $f_X(\bar{x}) > 0$, where f_X , the marginal probability density of X , is continuous at \bar{x} and for all ε , all $m \in \Omega$ satisfy

$$m(\bar{x}, \varepsilon) = \varepsilon$$

- Then, all the inverse functions, r , must satisfy

$$r(\varepsilon, \bar{x}) = \varepsilon$$

- Suppose r, \tilde{r} are any two such functions and for a strictly increasing s , and all ε, x

$$\tilde{r}(\varepsilon, x) = s(r(\varepsilon, x))$$

- Then, letting $x = \bar{x}$, it follows that for any t

$$t = \tilde{r}(t, \bar{x}) = s(r(t, \bar{x})) = s(t)$$

- Hence, s is the identity function.

- Clearly, if m^* is identified, so is F_ε^* , since for all e and any x

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

- In this expression, the first equality follows by the definition of F_{ε}^* , the second by the independence between ε and X , the third by the strict monotonicity of m^* in its last coordinate, and the last equality follows by the definition of Y and that of $F_{Y|X}$.

- It is also clear that if F_{ε}^* is specified, then m^* is identified, since from the above equation it follows that

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

- Imbens and Newey (2003) and Blundell and Powell (2003), for example, use a normalization that amounts to specifying ε to be $U(0, 1)$.

3.3.1. Identification of derivatives

- Rather than normalizing the set of functions, as above, we may ask what features can be identified without normalizations.
- It turns out that derivatives and discrete changes are identified.
- For the first result, let \bar{x} and \bar{y} denote particular values of, respectively, X and Y .
- Let $\bar{\varepsilon}$ denote the value of ε at which $\bar{y} = m^*(\bar{x}, \bar{\varepsilon})$.
- Assume that ε and X have differentiable densities, strictly positive at $\bar{\varepsilon}$ and \bar{x} , and that m^* is differentiable at $(\bar{x}, \bar{\varepsilon})$.
- Let Ω denote the set of all values that $\partial m^*(\bar{x}, \bar{\varepsilon}) / \partial x$ may attain.
- Then,

$$(3.c) \quad \psi^* = \partial m^*(\bar{x}, \bar{\varepsilon}) / \partial x \text{ is identified.}$$

- **Proof of (3.c):** We follow closely Matzkin (1999) and Chesher (2003).
- By independence between X and ε and the strict monotonicity of m ,

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

- Taking derivatives with respect to x , on both sides, we get that

$$0 = \frac{\partial F_{Y|X=\bar{x}}(t)}{\partial x} \Big|_{t=m^*(\bar{x}, \bar{\varepsilon})} + \frac{\partial F_{Y|X=\bar{x}}(t)}{\partial t} \Big|_{t=m^*(\bar{x}, \bar{\varepsilon})} \frac{\partial m^*(\bar{x}, \bar{\varepsilon})}{\partial x}$$

- Hence, the derivative

$$\frac{\partial m^*(\bar{x}, \bar{\varepsilon})}{\partial x} = - \left[\frac{\partial F_{Y|X=\bar{x}}(\bar{y})}{\partial y} \right]^{-1} \frac{\partial F_{Y|X=\bar{x}}(\bar{y})}{\partial x}$$

is uniquely derived from the distribution $F_{Y,X}$ of the observable variables. ■

3.3.2. Identification of finite changes

- Finite changes can also be identified.
- Fix again the value of (Y, X) at (\bar{y}, \bar{x}) , and let again $\bar{\varepsilon}$ be such that $\bar{y} = m^*(\bar{x}, \bar{\varepsilon})$.
- We are interested in the value of $y' - \bar{y}$ where $y' = m^*(x', \bar{\varepsilon})$.
- This is the causal effect on Y of changing the value of X from \bar{x} to x' , while leaving the value of the unobservable variable, ε , unchanged.
- Assume that the probability density f_X^* has a continuous extension and is strictly positive at \bar{x} and x' , and that the density of ε is strictly positive at $\bar{\varepsilon}$.
- Let Ω denote the set of all values that $y' - \bar{y}$ may attain.
- Then,

$$(3.d) \quad \psi^* = m^*(x', \bar{\varepsilon}) - m^*(\bar{x}, \bar{\varepsilon}) \text{ is identified.}$$

3.3.2. Identification of finite changes

- **Proof of (3.d):** The independence between X and ε and the strict monotonicity of m imply that

$$F_{\varepsilon}(\bar{\varepsilon}) = F_{Y|X=\bar{x}}(m^*(\bar{x}, \bar{\varepsilon}))$$

and, similarly, that

$$F_{\varepsilon}(\bar{\varepsilon}) = F_{Y|X=x'}(m^*(x', \bar{\varepsilon}))$$

- The strict monotonicity of $F_{Y|X=x'}$ then implies that

$$\begin{aligned}y' - \bar{y} &= m^*(x', \varepsilon^*) - \bar{y} \\&= F_{Y|X=x'}^{-1}(F_\varepsilon(\varepsilon^*)) - \bar{y} \\&= F_{Y|X=x'}^{-1}(F_{Y|X=\bar{x}}(m^*(\bar{x}, \bar{\varepsilon}))) - \bar{y} \\&= F_{Y|X=x'}^{-1}(F_{Y|X=\bar{x}}(\bar{y})) - \bar{y}\end{aligned}$$

- Hence, the change in the value of Y when X is changed from \bar{x} to x' is identified. ■

3.3.3. Identification in triangular systems

- In a model with a nonadditive, unobserved efficiency variable, Pakes and Olley (1996) used the strict monotonicity between investment and the unobserved index variable, conditional on observable age and capital stock of the firm, to express the unobserved efficiency index in terms of the observables age, capital stock, and investment.
- In a similar spirit, Chesher (2003) derived expressions for unobserved variables from conditional distributions, and use them to derive expressions for the derivatives of functions in a triangular system of equations with nonadditive random terms.
- Chesher used a local independence assumption.
- We will analyze here a special case of Chesher's model where the independence restrictions are stronger.

- To provide an example, suppose that the model of consumer demand is

$$Y = m(p, I, \varepsilon, \eta)$$

where ε and η are unobservable variables and m is strictly increasing in η .

- Suppose that I is determined by ε and an observable variable Z , according to a function \tilde{r} , strictly increasing in ε :

$$I = \tilde{r}(Z, \varepsilon)$$

- Assume that Z is distributed independently of (ε, η) . For simplicity, assume full support for all variables and differentiability for all functions.
- Then,

$$(3.e) \quad \frac{\partial m(p, I, \varepsilon, \eta)}{\partial I} \text{ can be identified}$$

- **Proof of (3.e):** Letting r denote the inverse of \tilde{r} with respect to ε and substituting in the demand function, we have that

$$Y = m(p, I, r(Z, I), \eta)$$

- Let

$$v(p, I, Z, \eta) = m(p, I, r(Z, I), \eta)$$

- Note that

$$\begin{aligned} \frac{\partial v(p, I, Z, \eta)}{\partial I} &= \frac{\partial m(p, I, r(Z, I), \eta)}{\partial I} \\ &+ \frac{\partial m(p, I, r(Z, I), \eta)}{\partial \varepsilon} \frac{\partial r(Z, I)}{\partial I} \end{aligned}$$

and

$$\frac{\partial v(p, I, Z, \eta)}{\partial Z} = \frac{\partial m(p, I, r(Z, I), \eta)}{\partial \varepsilon} \frac{\partial r(Z, I)}{\partial Z}$$

- Hence,

$$\frac{\partial m(p, l, \varepsilon, \eta)}{\partial l} \Big|_{\varepsilon=r(Z, l)} = \frac{\partial v(p, l, Z, \eta)}{\partial l} - \frac{\partial v(p, l, Z, \eta)}{\partial Z} \begin{bmatrix} \frac{\partial r(Z, l)}{\partial l} \\ \frac{\partial r(Z, l)}{\partial Z} \end{bmatrix}$$

- This implies that, if we know the functions v and r , we can identify the derivative of m with respect to I , at particular values of ε and δ .
- But, the models

$$I = \tilde{r}(Z, \varepsilon)$$

and

$$Y = v(p, I, Z, \eta)$$

are just the Independent Nonadditive Model, when ε and Z are independently distributed, and when (p, I, Z) and η are also independently distributed.

- Hence, the derivatives of \tilde{r} and of v are identified from the distribution of, respectively, (I, Z) and (Y, p, I, Z) .
- In particular, using the results in the previous section, it immediately follows that

$$\frac{\partial v(p, I, Z, \eta)}{\partial I} = - \left[\frac{\partial F_{Y|I,Z}(y^*)}{\partial y} \right]^{-1} \frac{\partial F_{Y|I,Z}(y^*)}{\partial I}$$

and

$$\frac{\partial v(p, I, Z, \eta)}{\partial Z} = - \left[\frac{\partial F_{Y|I,Z}(y^*)}{\partial y} \right]^{-1} \frac{\partial F_{Y|I,Z}(y^*)}{\partial Z}$$

at y^* such that $y^* = m(I, Z, \eta)$.

- Differentiating the expression

$$F_{\varepsilon}(r(Z, I)) = F_{Y|I,Z}(y)$$

which can be shown to be equivalent to the expression

$$F_{\varepsilon}(\varepsilon) = F_{Y|I,Z}(\tilde{r}(I, Z, \varepsilon))$$

we get, similarly, that

$$\frac{\partial r(Z, I)}{\partial I} = - \left[\frac{\partial F_{\varepsilon}(\varepsilon)}{\partial \varepsilon} \Big|_{\varepsilon=r(Z, I)} \right]^{-1} \frac{\partial F_{Y|I,Z}(y)}{\partial I}$$

and

$$\frac{\partial r(Z, I)}{\partial Z} = - \left[\frac{\partial F_{\varepsilon}(\varepsilon)}{\partial \varepsilon} \Big|_{\varepsilon=r(Z, I)} \right]^{-1} \frac{\partial F_{Y|I,Z}(y)}{\partial Z}$$

- Hence,

$$\frac{\partial m(p, l, \varepsilon, \eta)}{\partial l} = \left[\frac{\partial F_{Y|l,Z}(y^*)}{\partial y} \right]^{-1} \left[\frac{\partial F_{Y|l,Z}(y^*)}{\partial Z} \left[\frac{\partial F_{Y|l,Z}(y^*)}{\partial l} \right] - \frac{\partial F_{Y|l,Z}(y^*)}{\partial l} \right]$$

at $\varepsilon = r(l, Z)$ and $y^* = m(p, l, \varepsilon, \eta)$.

- Hence, using the variable Z we can identify the derivative of m with respect to l , leaving the value of ε fixed. ■

3.4. Identification in Nonadditive Index Models

- Consider the model,

$$Y = m^*(h^*(X), \varepsilon)$$

where Y denotes an observable dependent variable, ε denotes an unobservable explanatory variable whose support is R , X denotes a vector of observable explanatory variables that possesses support $\Xi = R^K$, X is such that the last coordinate, X_K , of X possesses an everywhere positive density conditional on the other coordinates of X , ε is distributed independently of X , $m^* : R^2 \rightarrow R$ is increasing in each coordinate, non-constant, and satisfies that for all t, t' ,

$$t < t' \Rightarrow \text{there exists } \varepsilon \text{ such that } m^*(t, \varepsilon) < m^*(t', \varepsilon)$$

and $h^* : \Xi \rightarrow R$ is continuous on Ξ and strictly increasing in its last coordinate.

- Assume that F_X is known.
- The model, S , is then characterized by the set of all triplets $\zeta = (\tilde{h}, \tilde{F}_\varepsilon, \tilde{m})$ such that \tilde{h} , \tilde{F}_ε , and \tilde{m} satisfy the assumptions that, respectively, h^* , F_ε^* , and m^* are assumed to satisfy.
- Let Ω denote the set composed of all first coordinates, \tilde{h} , of $(\tilde{h}, \tilde{F}_\varepsilon, \tilde{m}) \in S$.
- Let \circ denote the composition of two functions, so that $(g \circ \tilde{h})(t) = g(\tilde{h}(t))$.
- The following theorem was stated in Matzkin (1994).
- It's proof is a modification of the identification result in Han (1987) for semiparametric index models.

- **Theorem 3.2:** *In the model described above, two functions $h, \tilde{h} \in \Omega$ are observationally equivalent if and only if there exists a continuous, strictly increasing function $g : R \rightarrow R$ such that $\tilde{h} = g \circ h$.*

- **Proof:** Suppose that for all x , $\tilde{h}(x) = g(h(x))$.
- Then, letting $\tilde{m}(t, e) = m(g^{-1}(t), e)$, it follows that for all x, e

$$\begin{aligned}\tilde{m}(\tilde{h}(x), e) &= m(g^{-1}(g(h(x))), e) \\ &= m(h(x), e)\end{aligned}$$

- Hence, for any distribution, F_ε ,
 $F_{Y,X}(\cdot, \cdot; h, F_\varepsilon, m) = F_{Y,X}(\cdot, \cdot; \tilde{h}, F_\varepsilon, \tilde{m})$.
- It follows that h and \tilde{h} are observationally equivalent.

- On the other hand, suppose that there exist no strictly increasing, continuous g such that $\tilde{h} = g \circ h$, then, there must exist $x', x'' \in \Xi$ such that

$$h(x') < h(x'') \quad \text{and} \quad \tilde{h}(x') > \tilde{h}(x'')$$

- By the properties of any \tilde{m}, m , specified by the model, this implies that there exist $\varepsilon, \tilde{\varepsilon}$ such that

$$m(h(x'), \varepsilon) < m(h(x''), \varepsilon) \quad \text{and} \quad \tilde{m}(\tilde{h}(x'), \tilde{\varepsilon}) > \tilde{m}(\tilde{h}(x''), \tilde{\varepsilon})$$

- Let $F_\varepsilon, \tilde{F}_\varepsilon$ be any distributions that have support R .

- By independence between X and ε , the full support of ε , and the monotonicity of m and \tilde{m} , this implies that

$$\Pr \left\{ (e', e'') \mid \left(\tilde{m} \left(\tilde{h}(x'), e' \right) > \tilde{m} \left(\tilde{h}(x''), e'' \right) \right) \right\} \\ > \Pr \left\{ (e', e'') \mid \left(\tilde{m} \left(\tilde{h}(x'), e' \right) < \tilde{m} \left(\tilde{h}(x''), e'' \right) \right) \right\}$$

while

$$\Pr \left\{ (e', e'') \mid \left(m \left(h(x'), e' \right) > m \left(h(x''), e'' \right) \right) \right\} \\ < \Pr \left\{ (e', e'') \mid \left(m \left(h(x'), e' \right) < m \left(h(x''), e'' \right) \right) \right\}$$

- Hence, either

$$\Pr \left\{ (e', e'') \mid \left(\tilde{m} \left(\tilde{h}(x'), e' \right) < \tilde{m} \left(\tilde{h}(x''), e'' \right) \right) \right\} \\ \neq \Pr \left\{ (e', e'') \mid \left(m \left(h(x'), e' \right) < m \left(h(x''), e'' \right) \right) \right\}$$

or

$$\Pr \left\{ (e', e'') \mid \left(\tilde{m} \left(\tilde{h}(x'), e' \right) > \tilde{m} \left(\tilde{h}(x''), e'' \right) \right) \right\} \\ \neq \Pr \left\{ (e', e'') \mid \left(m \left(h(x'), e' \right) > m \left(h(x''), e'' \right) \right) \right\}$$

- Let $F_{Y,X}(\cdot; \tilde{h}, \tilde{F}_\varepsilon, \tilde{m})$ and $F_{Y,X}(\cdot; h, F_\varepsilon, m)$ denote the distributions generated by, respectively, $(\tilde{h}, \tilde{F}_\varepsilon, \tilde{m})$ and (h, F_ε, m) .
- Let Y' and Y'' denote the random variables that have, respectively, distributions $F_{Y|X=x'}$ and $F_{Y|X=x''}$.
- If any of the two inequalities above are satisfied, the probability of the event $Y' > Y''$ calculated using $F_{Y|X=x'}(\cdot, \cdot; \tilde{m}, \tilde{h}, \tilde{F}_\varepsilon)$ and $F_{Y|X=x''}(\cdot, \cdot; \tilde{m}, \tilde{h}, \tilde{F}_\varepsilon)$ will be different from the probability of the same event calculated using $F_{Y|X=x'}(\cdot, \cdot; m, h, F_\varepsilon)$ and $F_{Y|X=x''}(\cdot, \cdot; m, h, F_\varepsilon)$.

- By continuity of the functions, and the support conditions of X , this will still hold for all \tilde{x}' and \tilde{x}'' in neighborhoods, respectively, of x' and x'' , which have positive probability.
- Hence, $F_{Y,X}(\cdot, \cdot; \tilde{m}, \tilde{h}, \tilde{F}_\varepsilon) \neq F_{Y|X}(\cdot, \cdot; m, h, F_\varepsilon)$.
- It follows that h and \tilde{h} are not observationally equivalent. ■

- This result implies that if the restrict h^* to belong to a set of functions such that no two functions in the set are strictly increasing transformations of each other, then in that set h^* is identified.
- Matzkin (1994) describes several such set of functions.
- (See also Section 4.4.)

3.5. Identification in Simultaneous Equations Models

- Consider the simultaneous equations model, described in Section 2.2.1.5, where $Y \in R^G$ denotes a vector of observable dependent variables, $X \in R^K$ denotes a vector of observable explanatory variables, $\varepsilon \in R^G$ denotes a vector of unobservable explanatory variables, and the relationship between these vectors is specified by a function $r^* : R^G \times R^K \rightarrow R^G$ such that

$$\varepsilon = r^*(Y, X)$$

- The set S consisted of vectors of twice differentiable functions $r : R^G \times R^K \rightarrow R^G$ and twice differentiable, strictly increasing distributions $F_{\varepsilon, X} : R^G \times R^K \rightarrow R$ such that (i) for all $F_{\varepsilon, X}$, ε and X are distributed independently of each other (ii) for all r , and all y, x $|\partial r(y, x)/\partial y| > 0$, (iii) for all r and all x, ε , there exists a unique value of y such that $\varepsilon = r(y, x)$, and (iv) for all r , all $F_{\varepsilon, X}$, and all x , the distribution of Y given $X = x$, induced by r and $F_{\varepsilon|X=x}$ has support R^G .

- For any $(r, F_{\varepsilon, X}) \in S$, condition (iii) implies that there exists a function h such that for all ε, X ,

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

- This is the reduced form system of the structural equations system determined by r .
- We will let h^* denote the reduced form function determined by r^* .

- A special case of this model is the linear system of simultaneous equations, where for some invertible, $G \times G$ matrix A and some $G \times K$ matrix B ,

$$\varepsilon = AY + BX$$

- Premultiplication by $(A)^{-1}$ yields the reduced form system

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

where $\Pi = -(A)^{-1} B$ and $\nu = (A)^{-1} \varepsilon$.

- The identification of the true values, A^* , B^* , of the matrices A and B , and the distribution of ε has been the object of study in the works by Koopmans (1949), Koopmans, Rubin, and Leipnik (1950), and Fisher (1966), among others, and it is treated in most econometrics textbooks.

- The chapters by Hausman (1983) and Hsiao (1983) present the main known results.
- Assume that $E(\varepsilon) = 0$, and $Var(\varepsilon) = \Sigma^*$, an unknown matrix.
- Let W denote the variance of ν .
- Π and W can be identified from the distribution of the observable variables (Y, X) .
- The identification of any element of (A^*, B^*, Σ^*) is achieved when it can be uniquely recovered from Π and $Var(\nu)$.
- A priori restrictions on A^* , B^* , and Σ^* are typically used to determine the existence of a unique solution for any element of (A^*, B^*, Σ^*) .
- (See Fisher (1966).)

- In an analogous way, one can obtain necessary and sufficient conditions to uniquely recover r^* and F_ε^* from the distribution of the observable variables (Y, X) , when the system of structural equations is nonparametric.
- The question of identification is whether we can uniquely recover the density f_ε^* and the function r^* from the conditional densities $f_{Y|X=x}$.

- Following the definition of observational equivalence, we can state that two functions r, \tilde{r} satisfying the assumptions of the model are observationally equivalent iff there exist $f_\varepsilon, \tilde{f}_\varepsilon$ such that $(f_\varepsilon, r), (\tilde{f}_\varepsilon, \tilde{r}) \in S$ and for all y, x

$$(3.5.1) \quad f_{\tilde{\varepsilon}}(\tilde{r}(y, x)) \left| \frac{\partial \tilde{r}(y, x)}{\partial y} \right| = f_{\varepsilon}(r(y, x)) \left| \frac{\partial r(y, x)}{\partial y} \right|$$

- The function \tilde{r} can be expressed as a transformation of (ε, x) .
- To see this, define

$$g(\varepsilon, x) = \tilde{r}(h(x, \varepsilon), x)$$

- Since

$$\left| \frac{\partial g(\varepsilon, x)}{\partial \varepsilon} \right| = \left| \frac{\partial \tilde{r}(h(x, \varepsilon), x)}{\partial y} \right| \left| \frac{\partial h(x, \varepsilon)}{\partial \varepsilon} \right|$$

it follows that $|\partial g(\varepsilon, x) / \partial \varepsilon| > 0$.

- Let $\tilde{\varepsilon} = \tilde{r}(y, x)$.
- Since, conditional on x , h is invertible in ε and \tilde{r} is invertible in y , it follows that g is invertible in ε .
- Substituting in (3.5.1), we get that $(\tilde{r}, f_{\tilde{\varepsilon}}) \in S$ is observationally equivalent to $(r, f_{\varepsilon}) \in S$ iff for all ε, x

$$f_{\tilde{\varepsilon}}(g(\varepsilon, x)) \left| \frac{\partial g(\varepsilon, x)}{\partial \varepsilon} \right| = f_{\varepsilon}(\varepsilon)$$

- The following theorem provides conditions guaranteeing that a transformation g of ε does not generate an observable equivalent pair $(\tilde{r}, \tilde{f}_{\tilde{\varepsilon}}) \in S$ of a pair $(r, f_{\varepsilon}) \in S$

- **Theorem 3.3 (Matzkin (2005)):** Let $(r, f_\varepsilon) \in S$.
- Let $g(\varepsilon, x)$ be such that $\tilde{r}(y, x) = g(r(y, x), x)$ and $\tilde{\varepsilon} = g(\varepsilon, x)$ are such that $(\tilde{r}, f_{\tilde{\varepsilon}}) \in S$, where $f_{\tilde{\varepsilon}}$ denotes the marginal density of $\tilde{\varepsilon}$.
- If for some ε, x , the rank of the matrix

$$\begin{pmatrix} \left(\frac{\partial g(\varepsilon, x)}{\partial \varepsilon} \right)' & \frac{\partial \log f_\varepsilon(u)}{\partial \varepsilon} - \frac{\partial \log \left| \frac{\partial g(\varepsilon, x)}{\partial \varepsilon} \right|}{\partial \varepsilon} \\ \left(\frac{\partial g(\varepsilon, x)}{\partial x} \right)' & - \frac{\partial \log \left| \frac{\partial g(\varepsilon, x)}{\partial \varepsilon} \right|}{\partial x} \end{pmatrix}$$

is strictly larger than G , then, $(\tilde{r}, f_{\tilde{\varepsilon}})$ is not observationally equivalent to (r, f_ε) .

- Alternatively, we can express an identification theorem for the function r^* .

- **Theorem 3.4 (Matzkin (2005)):** Let $M \times \Gamma$ denote the set of pairs $(r, f_\varepsilon) \in S$.
- The function r^* is identified in M if $r^* \in M$ and for all $f_\varepsilon \in \Gamma$ and all $\tilde{r}, r \in M$ such that $\tilde{r} \neq r$, there exist y, x such that the rank of the matrix

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

is strictly larger than G, \dots

... where

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

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

- **Example 3.1:** As a very simple example, consider the simultaneous equations model, analyzed in Matzkin (2007c), where for some unknown function, g^* , and some parameter values β^*, γ^* ,

$$\begin{aligned}y_1 &= g^*(y_2) + \varepsilon_1 \\y_2 &= \beta^* y_1 + \gamma^* x + \varepsilon_2\end{aligned}$$

- Assume that $(\varepsilon_1, \varepsilon_2)$ has an everywhere positive, differentiable density $f_{\varepsilon_1, \varepsilon_2}^*$ such that for two, not necessarily known a-priori, values $(\bar{\varepsilon}_1, \bar{\varepsilon}_2)$ and $(\varepsilon_1'', \varepsilon_2'')$,

$$0 \neq \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}^*(\bar{\varepsilon}_1, \bar{\varepsilon}_2)}{\partial \varepsilon_1} \neq \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}^*(\varepsilon_1'', \varepsilon_2'')}{\partial \varepsilon_1} \neq 0$$

and

$$\frac{\partial \log f_{\varepsilon_1, \varepsilon_2}^*(\bar{\varepsilon}_1, \bar{\varepsilon}_2)}{\partial \varepsilon_2} = \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}^*(\varepsilon_1'', \varepsilon_2'')}{\partial \varepsilon_2} = 0$$

- The observable exogenous variable x is assumed to be distributed independently of $(\varepsilon_1, \varepsilon_2)$ and to possess support R . In this model

$$\varepsilon_1 = r_1^*(y_1, y_2, x) = y_1 - g^*(y_2)$$

$$\varepsilon_2 = r_2^*(y_1, y_2, x) = -\beta^* y_1 + y_2 - \gamma^* x$$

- The Jacobian determinant is

$$\left| \begin{pmatrix} 1 & -\frac{\partial g^*(y_2)}{\partial y_2} \\ -\beta^* & 1 \end{pmatrix} \right| = 1 - \beta^* \frac{\partial g^*(y_2)}{\partial y_2}$$

which will be positive as long as $1 > \beta^* \partial g^*(y_2)/\partial y_2$.

- Since the first element in the diagonal is positive, it follows by Gale and Nikaido (1965) that the function r^* is globally invertible if the condition $1 > \beta^* \partial g^*(y_2)/\partial y_2$ holds for every y_2 .
- Let r, \tilde{r} any two differentiable functions satisfying this condition and the other properties assumed about r^* .
- Suppose that at some y_2 , $\partial \tilde{g}(y_2)/\partial y_2 \neq \partial g(y_2)/\partial y_2$.
- Assume also that $\gamma \neq 0$ and $\tilde{\gamma} \neq 0$.
- Let $f_{\varepsilon_1, \varepsilon_2}$ denote any density satisfying the same properties that $f_{\varepsilon_1, \varepsilon_2}^*$ is assumed to satisfy.

- Denote by $(\varepsilon_1, \varepsilon_2)$ and $(\varepsilon'_1, \varepsilon'_2)$ the two points such that

$$0 \neq \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(\varepsilon_1, \varepsilon_2)}{\partial \varepsilon_1} \neq \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(\varepsilon'_1, \varepsilon'_2)}{\partial \varepsilon_1} \neq 0$$

and

$$\frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(\varepsilon_1, \varepsilon_2)}{\partial \varepsilon_2} = \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(\varepsilon'_1, \varepsilon'_2)}{\partial \varepsilon_2} = 0$$

- Define

$$a_1(y_1, y_2, x) = \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \varepsilon_1} - \beta \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \varepsilon_2},$$

$$a_2(y_1, y_2, x) = \left(\left(\frac{\partial^2 g(y_2)/\partial y_2^2}{(1 - \beta(\partial g(y_2)/\partial y_2))} \right) - \left(\frac{\partial^2 \tilde{g}(y_2)/\partial y_2^2}{(1 - \tilde{\beta}(\partial \tilde{g}(y_2)/\partial y_2))} \right) \right) - \left(\frac{\partial g(y_2)}{\partial y_2} \right) \left(\frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \varepsilon_1} \right) + \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \varepsilon_2}$$

and

$$a_3(y_1, y_2, x) = -\gamma \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \varepsilon_2}$$

- By Theorem 3.4, r and \tilde{r} will not be observationally equivalent if for all $f_{\varepsilon_1, \varepsilon_2}$ there exists (y_1, x) such that the rank of the matrix

$$A = \begin{pmatrix} 1 & -\tilde{\beta} & a_1(y_1, y_2, x) \\ -\frac{\partial \tilde{g}(y_2)}{\partial y_2} & 1 & a_2(y_1, y_2, x) \\ 0 & -\tilde{\gamma} & a_3(y_1, y_2, x) \end{pmatrix}$$

is 3.

- Let

$$a'_1(y_1, y_2, x) = (\tilde{\beta} - \beta) \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \varepsilon_2}$$

$$a'_2(y_1, y_2, x) = \left(\left(\frac{\partial^2 g(y_2)/\partial y_2^2}{(1 - \beta(\partial g(y_2)/\partial y_2))} \right) - \left(\frac{\partial^2 \tilde{g}(y_2)/\partial y_2^2}{(1 - \tilde{\beta}(\partial \tilde{g}(y_2)/\partial y_2))} \right) \right) + \left(\frac{\partial \tilde{g}(y_2)}{\partial y_2} - \frac{\partial g(y_2)}{\partial y_2} \right) \left(\frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \varepsilon_1} \right)$$

and

$$a'_3(y_1, y_2, x) = (\tilde{\gamma} - \gamma) \frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \varepsilon_2}$$

- Multiplying the first column of A by
 - $\partial \log f_{\varepsilon_1, \varepsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x) / \partial \varepsilon_1$ and adding it to the third column, and multiplying the second column by
 - $\partial \log f_{\varepsilon_1, \varepsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x) / \partial \varepsilon_2$ and adding it to the third column, we obtain the matrix

$$\begin{pmatrix} 1 & -\tilde{\beta} & a'_1(y_1, y_2, x) \\ -\frac{\partial \tilde{g}(y_2)}{\partial y_2} & 1 & a'_2(y_1, y_2, x) \\ 0 & -\tilde{\gamma} & a'_3(y_1, y_2, x) \end{pmatrix}$$

which has the same rank as A .

- By assumption, either

$$a'_2(y_1, y_2, x) = \left(\left(\frac{\partial^2 g(y_2)/\partial y_2^2}{(1 - \beta(\partial g(y_2)/\partial y_2))} \right) - \left(\frac{\partial^2 \tilde{g}(y_2)/\partial y_2^2}{(1 - \tilde{\beta}(\partial \tilde{g}(y_2)/\partial y_2))} \right) \right) + \left(\frac{\partial \tilde{g}(y_2)}{\partial y_2} - \frac{\partial g(y_2)}{\partial y_2} \right) \left(\frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(\varepsilon_1, \varepsilon_2)}{\partial \varepsilon_1} \right) \neq 0$$

or

$$a'_2(y_1, y_2, x) = \left(\left(\frac{\partial^2 g(y_2)/\partial y_2^2}{(1 - \beta(\partial g(y_2)/\partial y_2))} \right) - \left(\frac{\partial^2 \tilde{g}(y_2)/\partial y_2^2}{(1 - \tilde{\beta}(\partial \tilde{g}(y_2)/\partial y_2))} \right) \right) + \left(\frac{\partial \tilde{g}(y_2)}{\partial y_2} - \frac{\partial g(y_2)}{\partial y_2} \right) \left(\frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(\varepsilon'_1, \varepsilon'_2)}{\partial \varepsilon_1} \right) \neq 0$$

- Suppose the latter.
- Let $y_1 = g(y_2) + \varepsilon'_1$ and let $x = (-\beta y_1 + y_2 - \varepsilon'_2) / \gamma$.
- It then follows that

$$\frac{\partial \log f_{\varepsilon_1, \varepsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \varepsilon_2} = 0$$

- At such y_1, x , the above matrix becomes the rank 3 matrix

$$\begin{pmatrix} 1 & -\tilde{\beta} & 0 \\ -\frac{\partial \tilde{g}(y_2)}{\partial y_2} & 1 & a'_2(y_1, y_2, x) \\ 0 & -\tilde{\gamma} & 0 \end{pmatrix}$$

- Hence, derivatives of g^* are identified.

- **Example 3.2:** A similar example provides sufficient conditions for the identification of a utility function and its distribution, in a multidimensional version of the utility maximization problem described in Section 2.
- Let the utility function U^* for products $1, \dots, G + 1$, for a consumer with unobservable tastes $\varepsilon_1, \dots, \varepsilon_G$, be specified as:

$$U^*(y_1, \dots, y_{G+1}, \varepsilon_1, \dots, \varepsilon_G) = v^*(y_1, \dots, y_G) + \sum_{g=1}^G \varepsilon_g y_g + y_{G+1}$$

where v^* is a strictly monotone, strictly concave, twice differentiable function and where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_G)$ is distributed independently of (p, I) with a differentiable density that has known convex support.

- (This utility specification was studied in Brown and Calsamiglia (2004) in their development of tests for utility maximization; it is a slight modification of the specification used in Brown and Matzkin (1998) to analyze the identification of a distribution of utility functions from the distribution of demand.)
- Normalize the price of the $G + 1 - th$ commodity to equal 1.
- Maximization of U^* with respect to (y_1, \dots, y_{G+1}) subject to the budget constraint $\sum_{g=1}^G p_g y_g + y_{G+1} = I$ yields the first order conditions

$$\varepsilon_g = p_g - \partial v^*(y_1, \dots, y_G) / \partial y_g \quad g = 1, \dots, G$$

$$y_{G+1} = I - \sum_{g=1}^G p_g y_g$$

- Let $Dv^*(y)$ and $D^2v^*(y)$ denote, respectively, the gradient and Hessian of v^* at $y = (y_1, \dots, y_G)$.
- The first set of G equations represent a system of simultaneous equations with observable endogenous variables (y_1, \dots, y_G) and observable exogenous variables (p_1, \dots, p_G) .
- The strict concavity of v^* guarantee that for any (p_1, \dots, p_G) and $(\varepsilon_1, \dots, \varepsilon_G)$, a unique solution for (y_1, \dots, y_G) exists.
- Let W denote the set of functions v satisfying the same restrictions that v^* is assumed to satisfy.
- Let $\bar{\varepsilon}$ denote a given value of the vector ε .

- Let Γ denote the set of all densities f_ε of ε such that (i) f_ε is differentiable, (ii) $f_\varepsilon(\varepsilon) > 0$ on a neighborhood of radius δ around $\bar{\varepsilon}$, (iii) for all ε in the support of f_ε , $\partial \log(f_\varepsilon(\varepsilon))/\partial \varepsilon = 0$ iff $\varepsilon = \bar{\varepsilon}$, (iv) for all g , there exist two distinct values, ε' and ε'' , in the δ -neighborhood of $\bar{\varepsilon}$ such that $f_\varepsilon(\varepsilon'), f_\varepsilon(\varepsilon'') > 0$, $0 \neq \partial \log(f_\varepsilon(\varepsilon'))/\partial \varepsilon_g \neq \partial \log(f_\varepsilon(\varepsilon''))/\partial \varepsilon_g \neq 0$, and for $j \neq g$, $\partial \log(f_\varepsilon(\varepsilon'))/\partial \varepsilon_j = \partial \log(f_\varepsilon(\varepsilon''))/\partial \varepsilon_j = 0$.
- Suppose that W and the support of p is such for all y , for all $v \in W$, there exist a set of prices, Q , such that the density of p is uniformly bounded away from zero on Q and the range of $Dv(y) - p$, when considered as a function of p over Q , is the δ neighborhood of $\bar{\varepsilon}$.
- Then, if v, \tilde{v} belong to W and $D\tilde{v} \neq Dv$, there exist, for all $f_\varepsilon \in \Gamma$, values y, p such that the rank of the corresponding matrix in Theorem 3.4 is larger than G .
- (See Matzkin (2007a))

3.6. Identification in Discrete Choice Models

- Consider the discrete choice model described in Section 2.2.1.6, where a typical individual has to choose between $G + 1$ alternatives.
- Let $V_g(s, z_g, \omega)$ denote the utility for alternative g , where s denotes a vector of observable characteristics of the consumer, z_g denotes a vector of observable attributes of alternative g , and ω is an unobservable random vector.
- The vector of observable dependent variables is $y = (y_1, \dots, y_{G+1})$ defined by

$$y_g = \left\{ \begin{array}{ll} 1 & \text{if } V_g(s, z_g, \omega) > V_k(s, z_g, \omega) \text{ for all } k \neq g \\ 0 & \text{otherwise} \end{array} \right\}$$

- Let z denote the vector (z_1, \dots, z_{G+1}) .
- The conditional choice probability, for each $g = 1, \dots, G + 1$ is

$$\Pr(\{y_g = 1 \mid s, z\}) = \Pr(\{\omega \mid V_g(s, z_g, \omega) > V_k(s, z_k, \omega) \text{ for all } k \neq g\})$$

- Since the choice probabilities of each alternative depend only on the differences between the utilities of the alternatives, only those differences can be identified.
- Hence, for simplicity, we may specify $V_{G+1}(s, z_{G+1}, \omega)$ equal to 0 for all (s, z_{G+1}, ω) .
- Then,

$$\Pr(\{y_{G+1} = 1 \mid s, z\}) = \Pr(\{\omega \mid 0 > V_k(s, z_k, \omega) \text{ for all } k \neq G + 1\})$$

- (We assume that the probability of ties is zero.)

3.6.1. Subutilities additive in the unobservables

- The simplest case to analyze is when $\omega = (\omega_1, \dots, \omega_G)$, each V_g depends only on one coordinate, ω_g of ω , and ω_g is additive:

$$V_g(s, z_g, \omega) = v_g(s, z_g) + \omega_g,$$

where v_g is a nonparametric function.

- (Matzkin (1991a) studies identification in this model when the distribution of ω is specified parametrically. Matzkin (1992, 1993, 1994) extends some of those results for the case of nonparametric distributions.)
- Under the additivity assumption:

$$\Pr(\{y_{G+1} = 1 \mid s, z\}) = F_{\omega_1, \dots, \omega_G}(-v_1(s, z_1), \dots, -v_G(s, z_G))$$

where $F_{\omega_1, \dots, \omega_G}$ is the unknown distribution of $(\omega_1, \dots, \omega_G)$.

- This is of the form of a multiple index model, and it could therefore be analyzed using techniques for those models.

- Assume, for example, that each of the z_g vectors includes a coordinate $z_g^{(1)}$ with is such that

$$v_g(s, z_g^{(1)}, z_g^{(2)}) = z_g^{(1)} + m_g(s, z_g^{(2)})$$

where $z_g = (z_g^{(1)}, z_g^{(2)})$ and m_g is a nonparametric function.

- Then,

$$\Pr(\{y_{G+1} = 1 \mid s, z\}) = F_{\omega_1, \dots, \omega_G} \left(-z_1^{(1)} - m_1(s, z_1^{(2)}), \dots, -z_G^{(1)} - m_G(s, z_G^{(2)}) \right)$$

- Assume that $(\omega_1, \dots, \omega_G)$ is distributed independently of (s, z_1, \dots, z_G) .
- Let $(\bar{s}, \bar{z}^{(2)}) = (\bar{s}, \bar{z}_1^{(2)}, \dots, \bar{z}_G^{(2)})$ denote a particular value of $(s, z^{(2)})$.
- Assume that $z^{(1)} = (z_1^{(1)}, \dots, z_G^{(1)}) \in R^G$ possesses an everywhere positive density on R^G , conditional on $(\bar{s}, \bar{z}^{(2)}) = (\bar{s}, \bar{z}_1^{(2)}, \dots, \bar{z}_G^{(2)})$.
- Let $\alpha_g \in R$ and specify that for $g = 1, \dots, G$

$$m_g(\bar{s}, \bar{z}_g^{(2)}) = \alpha_g$$

- Then,

$$\Pr(\{y_{G+1} = 1 \mid \bar{s}, z^{(1)}, \bar{z}^{(2)}\}) = F_{\omega_1, \dots, \omega_G}(-z_g^{(1)} - \alpha_g, \dots, -z_g^{(1)} - \alpha_G),$$

which shows that $F_{\omega_1, \dots, \omega_G}$ can be recovered from the choice probabilities, evaluated at appropriate values of $(s, z^{(1)}, z^{(2)})$.

- In an influential paper, Lewbel (2000) shows that the requirement that $(\omega_1, \dots, \omega_G)$ be independent of (s, z) is not needed for identification of $F_{\omega_1, \dots, \omega_G}$.
- It suffices that $(\omega_1, \dots, \omega_G)$ be independent of $z^{(1)}$ conditional on $(s, z^{(2)})$, in addition to the large support condition on $z^{(1)}$.
- Since the work of Lewbel (2000), the vector $z^{(1)}$ has been called a "special regressor".
- Its identification force has been extended to many models other than discrete choice models.

3.6.2. Subutilities nonadditive in the unobservables

- Applying Lewbel's special regressor technique, one can analyze models with nonadditive unobservables, as described in Matzkin (2005b).
- Suppose that each V_g is specified as:

$$V_g(s, z_g^{(1)}, z_g^{(2)}, \omega) = z_g^{(1)} + v_g(s, z_g^{(2)}, \omega)$$

where v_g is a nonparametric function.

- Assume that ω is distributed independently of (s, z) .
- Define Υ_g for each g by

$$\Upsilon_g = v_g(s, z_g^{(2)}, \omega)$$

- Since ω is distributed independently of (s, z) , $(\Upsilon_1, \dots, \Upsilon_G)$ is distributed independently of $z^{(1)}$, conditional on $(s, z^{(2)})$.
- Hence, using the arguments in Lewbel (2000), one can recover the distribution of $(\Upsilon_1, \dots, \Upsilon_G)$ given $(s, z^{(2)})$.
- From this distribution, one can identify the functions v_1, \dots, v_G and the distribution of $(\omega_1, \dots, \omega_G)$ in the system

$$\begin{aligned}\Upsilon_1 &= v_1(s, z_1^{(2)}, \omega_1, \dots, \omega_G) \\ \Upsilon_2 &= v_2(s, z_2^{(2)}, \omega_1, \dots, \omega_G) \\ &\quad \cdot \quad \cdot \\ \Upsilon_G &= v_G(s, z_G^{(2)}, \omega_1, \dots, \omega_G)\end{aligned}$$

using the results in Matzkin (2005).

- In particular, assume that, given $(s, z^{(2)})$, the system of functions (v_1, \dots, v_G) is invertible in ω .
- Then, it can be equivalently expressed as

$$\omega = r(\Upsilon, s, z^{(2)})$$

where ω is the vector $(\omega_1, \dots, \omega_G)'$ and $\Upsilon = (\Upsilon_1, \dots, \Upsilon_G)$.

- This has the same structure as considered in the previous sections.
- (See Matzkin (2007a) for more detail.)
- Unobservable vectors of dimension larger than G can be dealt with making use of additional functional restrictions and conditional independence assumptions.
- (See the Appendix in Matzkin (2003).)

Ways of achieving identification

- When a feature of interest is not identified, one may proceed in different ways to achieve identification.
- One may augment the model, incorporating more observable variables.
- One may impose further restrictions on either the functions, or the distributions, or both.
- The analysis of observational equivalence together with economic theory can often be used to determine appropriate restrictions.
- In this section, we describe examples of some of the techniques that have been developed, following one or more of these approaches.
- The emphasis will be in showing how one can recover particular features, once objects such as conditional distributions and conditional expectations are identified.

Conditional independence

- A common situation encountered in econometric models is where the unobservable variables affecting the value of an outcome variable are not distributed independently of the observed explanatory variables.
- Without additional information, identifying the causal effect of the observable explanatory variables on the outcome variable is typically not possible in such a situation.
- Usually, the additional information involves variables and restrictions guaranteeing some exogenous variation on the value of the explanatory variable.
- The leading procedures to achieve this are based on conditional independence methods and instrumental variable methods.

- In the first set of procedures, independence between the unobservable and observable explanatory variables in a model is achieved after conditioning on some event, some function, or some value of an external variable or function.
- The second set of procedures usually derives identification from an independence condition between the unobservable and an external variable (an instrument) or function.
- In this subsection, we will deal with conditional independence.
- In Subsection 4.2, we will deal with instrumental variables.

4.1.1. Identification of functions and distributions in a nonadditive model using conditional independence

- Consider the nonadditive model

$$Y_1 = m_1(X, \varepsilon_1)$$

where ε and X are not independently distributed and m is strictly increasing in ε .

- A standard example (see Chesher (2003) and Imbens and Newey (2003)) is where Y_1 denotes earnings, X denotes years of education, and ε denotes the effect of unobservable explanatory variables, which includes unobserved ability.
- Since X is determined as a function of ε , these variables are not independently distributed.

- Suppose, however, that some variable W is available, such that for some function m_2 and some ε_2 ,

$$X = m_2(W, \varepsilon_2)$$

- Denoting X by Y_2 , the system of the two above equations is a triangular system.
- Imbens and Newey (2003) developed identification results for this system when W is observable and independent of $(\varepsilon_1, \varepsilon_2)$.
- Chesher (2003) considered local independence conditions for identification of local derivatives.

- Matzkin (2004) studied identification when ε_1 and ε_2 are independent, conditional on either a particular value or all possible values of W .
- (A footnote in Chesher (2003) also discusses independence restrictions on the unobservables as a source of identification.)
- When W is independent of $(\varepsilon_1, \varepsilon_2)$, independence between ε_1 and X can be determined conditional on the unobservable ε_2 .
- When ε_1 and ε_2 are independent conditional on W , independence between ε_1 and X can be determined conditional on the observable W .

- The following theorem, in Matzkin (2004), provides insight into the sources of identification
- **Theorem 4.1 (Equivalence Theorem (Matzkin (2004)):**
Consider the model $Y_1 = m_1(X, \varepsilon_1)$.
- *Suppose that m_1 is strictly increasing in ε_1 , and that for all values w of W , the conditional distribution, $F_{X, \varepsilon | W=w}$, of (X, ε) given $W = w$ is strictly increasing.*
- *Then, the following statements are equivalent:*
- *(i) There exists a strictly increasing function $m_2(W, \cdot)$ and an unobservable random term ε_2 such that $X = m_2(W, \varepsilon_2)$ and ε_2 is independent of (W, ε_1) .*
- *(ii) There exists a strictly increasing function $r(W, \cdot)$ and an unobservable random term δ such that $\varepsilon_1 = r(W, \delta)$, δ is independent of (X, W) .*
- *(iii) ε_1 is independent of X , conditional on W .*

- Consider the Nonadditive Model

$$Y_1 = m_1(X, \varepsilon_1)$$

- To be able to identify m_1 , we need to observe independent variation in each coordinate of m .
- The theorem considers three different representations of the model:

$$\begin{aligned} Y &= m_1(m_2(W, \varepsilon_2), \varepsilon_1) \\ &= m_1(m_2(W, \varepsilon_2), r(W, \delta)) \\ &= m_1(X, r(W, \delta)) \end{aligned}$$

- From the first expression, it follows that if ε_1 and ε_2 are independent conditional on at least one value \bar{w} of W , then we will be able to observe events where, conditional on W , each coordinate of m_1 achieves values independently of the other coordinates of m_1 .
- From the third expression, it follows that if δ is independent of X conditional on at least one value \bar{w} of W , then, again each coordinate of m_1 will achieve values independently of the other coordinates of m_1 , when conditioning on at least one value of W .

- The second expression provides the same result, when we can establish that δ and ε_2 are independent, conditional on at least one value \bar{w} of W .
- The equivalence theorem above states that as long as we show that the conditions for one of these representations are satisfied, then the conditions for the other representations also hold.
- The above theorem also holds when W is unobservable, ε_2 is observable, and ε_2 is distributed independently of (ε_1, W) .
- In such case, that (i) implies (iii) is shown in Imbens and Newey (2003) as follows: The restriction that ε_2 is independent of (W, ε_1) implies that, conditional on W , ε_2 and ε_1 are independent.
- Since conditional on W , X is a function of ε_2 , and ε_2 is independent of ε_1 , it follows that conditional on W , X is independent of ε_1 .

- The local condition, that conditional on $W = \bar{w}$, ε_1 and ε_2 are independent, can be shown to imply, under some additional assumptions, that m_1 and the distribution of (X, ε_1) can both be identified, up to a normalization on the distribution of ε_1 given $W = \bar{w}$.

- In particular, Matzkin (2004) shows that if m_1 is strictly increasing in ε_1 , $F_{\varepsilon_1, X|W=\bar{w}}$ is strictly increasing in (ε_1, X) , for each x , $F_{\varepsilon_1|(X,W)=(x,\bar{w})}$ is strictly increasing in ε_1 , and if there exists a function m_2 and an unobservable ε_2 such that $X = m_2(W, \varepsilon_2)$, m_2 is strictly increasing in ε_2 when $W = \bar{w}$, and ε_1 is independent of ε_2 conditional on $W = \bar{w}$, then for all x, e

$$(4.a) \quad m(x, e) = F_{Y|(X,W)=(x,\bar{w})}^{-1} (F_{\varepsilon_1|W=\bar{w}}(e))$$

and

$$F_{\varepsilon_1|X=x}(e) = F_{Y|X=x} \left(F_{Y|(X,W)=(x,\bar{w})}^{-1} (F_{\varepsilon_1|W=\bar{w}}(e)) \right)$$

- Matzkin (2004) describes several examples where economic theory implies the conditional exogeneity of the unobservable ε_2 , for particular variables W .

- **Proof of (4.a):** Let x be given and let e_2 denote the value of ε_2 such that $x = m_2(\bar{w}, e_2)$.
- By conditional independence and strict monotonicity

$$\begin{aligned} \Pr(\varepsilon_1 \leq e | W = \bar{w}) &= \Pr(\varepsilon_1 \leq e | \varepsilon_2 = e_2, W = \bar{w}) \\ &= \Pr(m_1(X, \varepsilon_1) \\ &\leq m_1(x, e) | X = m_1(\bar{w}, e_2), W = \bar{w}) \\ &= F_{Y_1|X=x, W=\bar{w}}(m_1(x, e)) \end{aligned}$$

- Hence,

$$m_1(x, e) = F_{Y_1|X=x, W=\bar{w}}^{-1}(F_{\varepsilon_1|W=\bar{w}}(e))$$

- Since

$$F_{\varepsilon_1|X=x} = F_{Y|X=x}(m_1(x, e))$$

it follows that

$$F_{\varepsilon_1|X=x} = F_{Y|X=x} \left(F_{Y_1|X=x, W=\bar{w}}^{-1}(F_{\varepsilon_1|W=\bar{w}}(e)) \right)$$



- As with the case where X and ε_1 are independently distributed, identification of derivatives of m_1 with respect to X does not require additional normalizations.
- Altonji and Matzkin (2001) present the following result (see also Altonji and Ichimura (2000)).

Identification of average derivatives in a nonadditive model using conditional independence

- Consider the nonseparable model

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

where no particular assumptions are made regarding monotonicity of m .

- Let $\varepsilon = (\varepsilon_1, \dots, \varepsilon_J)$.
- Assume that m and the density $f_{\varepsilon|X}$ are differentiable with respect to X in a neighborhood of a value x of X , that $f_{\varepsilon|X}$ is everywhere positive in ε and the marginal density f_X is strictly positive on a neighborhood of x .

- Assuming that the integral and all the terms inside the integral exist, suppose that we wanted to recover the average derivative

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

using a conditioning vector of variables W .

- Altonji and Matzkin (2001, 2005) show that if ε is independent of X conditional on W , then

(4.b) $\beta(x)$ can be recovered from the distribution of the observable variables

- **Proof of (4.b):** Since for all e, x, w ,

$$f_{\varepsilon|W=w, X=x}(e) = f_{\varepsilon|W=w}(e)$$

one has that

$$\frac{\partial f_{\varepsilon|W=w, X=x}(e)}{\partial x} = 0$$

- Let $E[Y|W = w, X = x]$ denote the conditional expectation of Y given $(W = w, X = x)$.

- Then,

$$\begin{aligned}
 & \int \frac{\partial E[Y|W = w, X = x]}{\partial x} f_{W|X=x}(w) dw \\
 = & \int \frac{\partial}{\partial x} \int m(x, \varepsilon) f_{\varepsilon|W=w, X=x}(\varepsilon) \frac{f_{W, X}(w, x)}{f_X(x)} d\varepsilon dw \\
 = & \int \left[\frac{\partial}{\partial x} \int m(x, \varepsilon) f_{\varepsilon|W=w, X=x}(\varepsilon) d\varepsilon \right] \frac{f_{W, X}(w, x)}{f_X(x)} dw \\
 = & \int \left[\int \frac{\partial m(x, \varepsilon)}{\partial x} f_{\varepsilon|W=w, X=x}(\varepsilon) d\varepsilon + \int m(x, \varepsilon) \frac{\partial f_{\varepsilon|W=w, X=x}(\varepsilon)}{\partial x} d\varepsilon \right] \frac{f_{W, X}(w, x)}{f_X(x)} dw \\
 = & \int \left[\int \frac{\partial m(x, \varepsilon)}{\partial x} f_{\varepsilon|W=w, X=x}(\varepsilon) d\varepsilon \right] \frac{f_{W, X}(w, x)}{f_X(x)} dw \\
 = & \int \int \frac{\partial m(x, \varepsilon)}{\partial x} \frac{f_{\varepsilon, W, X}(\varepsilon, w, x)}{f_{W, X}(w, x)} \frac{f_{W, X}(w, x)}{f_X(x)} d\varepsilon dw \\
 = & \int \int \frac{\partial m(x, \varepsilon)}{\partial x} \frac{f_{\varepsilon, W, X}(\varepsilon, w, x)}{f_X(x)} dw d\varepsilon \\
 = & \int \frac{\partial m(x, \varepsilon)}{\partial x} \frac{f_{\varepsilon, X}(\varepsilon, x)}{f_X(x)} d\varepsilon \\
 = & \int \frac{\partial m(x, \varepsilon)}{\partial x} f_{\varepsilon|X=x}(\varepsilon) d\varepsilon \\
 = & \beta(x)
 \end{aligned}$$

- Since $E[Y|W = w, X = x]$ and $f_{W|X}$ can be recovered from the distribution of (Y, W, X) , $\beta(x)$ can also be recovered from it. ■
- Many other functions, average derivatives, and other functions can be derived and shown to be identified in the nonadditive model $Y_1 = m_1(X, \varepsilon_1)$.

- Blundell and Powell (2003) consider identification and estimation of the "average structural function", defined for $X = x$ as

$$G(x) = \int m_1(x, e) f_{\varepsilon_1}(e) de$$

- Blundell and Powell (2003) assumed the existence of a random vector

$$\nu = \nu(y, x, w)$$

which is identified and estimable, and it is such that the distribution of ε_1 conditional on (X, W) is the same as the distribution of ε_1 conditional on (X, ν) , which is the same as the distribution of ε_1 conditional on ν .

- The average structural function is then obtained from the distribution of (Y, X, ν) as:

$$G(x) = \int E(Y|X, \nu) f_{\nu}(\nu) d\nu$$

- This follows because

$$\begin{aligned} G(x) &= \int m_1(x, e) f_{\varepsilon_1}(e) de \\ &= \int \left[\int m_1(x, e) f_{\varepsilon_1|\nu}(e) de \right] f_{\nu}(\nu) d\nu \\ &= \int [E(Y|X, \nu)] f_{\nu}(\nu) d\nu \end{aligned}$$

- Imbens and Newey (2003) consider identification of the "quantile structural function", defined for $\tau \in (0, 1)$ and all x as

$$m_1(x, q_{\varepsilon_1}(\tau))$$

where $q_{\varepsilon_1}(\tau)$ is the τ -quantile of the distribution of ε_1 .

- Letting ν be such that ε_1 is independent of X conditional on ν , they obtain the following expression for the inverse $m_1^{-1}(x, y)$ of m_1 with respect to $q_{\varepsilon_1}(\tau)$:

$$\begin{aligned} m_1^{-1}(x, y) &= \Pr(m_1(x, q_{\varepsilon_1}(\tau)) \leq y) \\ &= \int \Pr(Y \leq y | \nu) f_\nu(\nu) d\nu \\ &= \int \Pr(Y \leq y | X = x, \nu) f_\nu(\nu) d\nu \end{aligned}$$

- For the average derivative, Imbens and Newey (2003) use the fact that, under conditional independence

$$\begin{aligned}\delta &= E \left[\frac{\partial m_1(x, \varepsilon_1)}{\partial x} \right] \\ &= E \left[\int \frac{\partial m_1(x, e)}{\partial x} f_{\varepsilon_1|X=x, \nu}(e) de \right] \\ &= E \left[\int \frac{\partial m_1(x, e)}{\partial x} f_{\varepsilon_1|X=x, \nu}(e) de \right] \\ &= E \left[\frac{\partial}{\partial x} E(Y|X = x, \nu) \right]\end{aligned}$$

Marginal independence

- In many situations, such as in models with simultaneity, establishing conditional independence between the unobservable and observable explanatory variables that determine the value of an outcome variable may require undesirable strong assumptions (see Blundell and Matzkin (2007)).
- A variable that is independent of the unobservable variables, and not independent of the observable variables may be used in such and other situations.
- In the model

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

where X is not distributed independently of ε , an instrument is a variable, Z , that is distributed independently of ε and is not distributed independently of X .

4.2.1. Instrumental variables in nonadditive models

- Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007), and Matzkin (2004, 2005b) consider identification of nonadditive models using instruments.
- Chernozhukov, Imbens, and Newey (2007)'s model is

$$Y = m(X, Z_1, \varepsilon)$$

where X is a vector of observable variables that is not distributed independently of ε , m is strictly increasing in ε , $Z = (Z_1, Z_2)$ is an observable vector that is distributed independently of ε , and the density of ε is everywhere positive.

- Since the distribution of ε and m are not jointly identified, one may normalize the marginal distribution of ε to be $U(0, 1)$.

- Independence between ε and Z imply that for each $\tau \in (0, 1)$

$$\begin{aligned}\tau &= E[1(\varepsilon < \tau)] = E[1(\varepsilon < \tau) | Z] \\ &= E[E[1(\varepsilon < \tau) | W, Z] | Z] \\ &= E[E[1(m(W, \varepsilon) < m(W, \tau)) | W, Z] | Z] \\ &= E[1(Y < m(W, \tau)) | Z]\end{aligned}$$

- Define $\rho(Y, W, \tau, m) = 1(Y < m(W, \tau)) - \tau$.
- Then, the above defines a conditional moment restriction

$$E[\rho(Y, W, \tau, m) | Z] = 0$$

- The following theorem provides sufficient conditions for local identification, in the sense of Rothenberg (1971), of $\rho(Y, W, \tau, m)$.
- **Theorem 4.2 (Chernozhukov, Imbens, and Newey (2004)):** *Suppose that Y is continuously distributed conditional on X and Z with density $f(y|x, z)$, and that there exists $C > 0$ such that*

$$|f(y|x, z) - f(\tilde{y}|x, z)| \leq C |y - \tilde{y}|$$

and for $D(V) = f(m(W, \tau) | W, Z)$, $E[D(V) \Delta(V) | Z] = 0$ implies $\Delta(V) = 0$ then $m(W, \tau)$ is locally identified.

- In simultaneous equations, of the type considered in previous sections, an observed or identified exogenous variable that is excluded from one equation may be used as an instrument for that equation.
- Consider, for example, the simultaneous equation model

$$\begin{aligned} Y_1 &= m_1(Y_2, \varepsilon_1) \\ Y_2 &= m_2(Y_1, X, \varepsilon_2) \end{aligned}$$

where X is distributed independently of $(\varepsilon_1, \varepsilon_2)$.

- Matzkin (2007b) establishes restrictions on the functions m_1 and m_2 and on the distribution of $(\varepsilon_1, \varepsilon_2, X)$ under which

$$\left[\frac{\partial r_1(y_1, y_2)}{\partial y_2} \right]^{-1} \left[\frac{\partial r_1(y_1, y_2)}{\partial y_1} \right]$$

can be expressed as a function of the values of $f_{Y_1, Y_2, X}$ at $(Y_1, Y_2) = (y_1, y_2)$ and particular values of X .

4.2.2. Unobservable Instruments

- Matzkin (2004) considers the use of *unobservable* instruments to identify nonadditive models.
- These are variables that are known to be distributed independently of unobservable random terms in an equation of interest, but are themselves unobservable.
- This is in the spirit of Fisher (1966), who developed an extensive set of conditions on the unobservables in linear systems of simultaneous equations that provide identification.
- The method is also related to the one in Hausman and Taylor (1983).

- Matzkin (2004) considers the model

$$Y_1 = m(Y_2, X, \varepsilon)$$

with m strictly increasing in ε and ε distributed independently of X .

- She assumes that a second equation,

$$Y_2 = g(Y_1, \eta)$$

is identified, and that the unobservables η and ε are independently distributed.

- The identification of the function g in general will require imposing additional restrictions.
- If, for example, g were specified to be a linear function and one assumed that $E[\eta|X] = 0$, then identification of g would follow by standard results.
- If g were nonparametric and additive in η , then, under the assumption that $E[\eta|X] = 0$ one could identify it using the methods in Newey and Powell (1989, 2003), Darolles, Florens, and Renault (2000), or Hall and Horowitz (2003).
- Suppose that g is identified.
- Matzkin (2004) proposes a pointwise direct identification of the function m .

- The argument proceeds by using η to estimate the reduced form equations

$$Y_1 = r_1(X, \eta, \varepsilon)$$

$$Y_2 = r_2(X, \eta, \varepsilon)$$

- Under the assumption that ε is independent of (X, η) , these equations are identified, using the arguments in 3.3.
- These equations are next used to identify m .
- To see this, suppose that we wanted to identify the value of m at a particular value (y_2, x, e) .

- Let η^* denote the value of η that solves the equation

$$y_2 = r_2(x, \eta^*, e)$$

- Let $y_1^* = r_1(x, \eta^*, e)$.
- It then follows by the definition of m and of the functions r_1 and r_2 that

$$\begin{aligned} m(y_2, x, e) &= m(r_2(x, \eta^*, e), x, e) \\ &= r_1(x, \eta^*, e) \\ &= y_1^* \end{aligned}$$

- Hence, one can recover the function m .

4.2.3. Instrumental variables in additive models

- In additive models, the requirement that $Z = (Z_1, Z_2)$ be independent of ε_1 may be weakened to a conditional mean independence.
- Newey and Powell (1989, 2003), Darolles, Florens, and Renault (2000), Ai and Chen (2003), and Hall and Horowitz (2005) considered the model

$$Y = m(X, Z_1) + \varepsilon$$

where $E[\varepsilon|X] \neq 0$.

- They assumed the existence of an instrument, Z , satisfying

$$E[\varepsilon|Z_1, Z_2] = 0$$

- Using the definition of ε , this yields the equation

$$\begin{aligned} E[Y|Z_1 = z_1, Z_2 = z_2] &= E[m(X, z_1) | Z_1 = z_1, Z_2 = z_2] \\ &= \int m(x, z_1) f_{X|Z_1=z_1, Z_2=z_2}(x) dx \end{aligned}$$

- Since the "reduced form" $E[Y|Z_1, Z_2]$ is identified from the distribution of (Y, Z_1, Z_2) and $f_{X|Z_1=z_1, Z_2=z_2}(x)$ is identified from the distribution of (X, Z) , the only unknown in the above integral equation is $m(x, z_1)$.
- Newey and Powell (2003) provided conditions characterizing the identification of the function m solely from the above integral equation.

- **Theorem 4.3 (Newey and Powell (2003)):** *Suppose that $Y = m(X, Z_1) + \varepsilon$ and $E[\varepsilon|Z_1, Z_2] = 0$.*
- *Then, m is identified if and only if for all functions $\delta(x, z_1)$ with finite expectation, $E[\delta(x, z_1)|Z = z] = 0$ implies that $\delta(x, z_1) = 0$.*
- Das (2004) and Newey and Powell (2003) considered identification of this model when the endogenous variables are discrete.
- To state the result presented in Newey and Powell (2003), assume that both X and Z_2 are discrete.
- Denote the support of X and Z_2 by, respectively, $\{x_1, \dots, x_S\}$ and $\{z_{21}, \dots, z_{2T}\}$.
- Let $P(z_1)$ denote the $S \times T$ matrix whose ij -th elements is $\Pr(X = x_j | Z_1 = z_1, Z_2 = z_{2j})$.

- **Theorem 4.4** (Newey and Powell (2003)): *Suppose that $Y = m(X, Z_1) + \varepsilon$, $E[\varepsilon|Z_1, Z_2] = 0$, and X and Z_2 have finite support.*
- *Then, $m(x, z_1)$ is identified if and only if $\Pr[\text{rank}(P(z_1)) = s] = 1$.*

4.2.4. Instrumental variables in additive models with measurement error

- A common situation where an observable explanatory variable is not independent of the unobserved explanatory variable is when the observed explanatory variable is an imperfect measurement of the true explanatory variable, which is unobserved.
- For this situation, Schennach (2007) established identification of an additive model using instrumental variables.
- She considered the model

$$\begin{aligned}Y &= m(X^*) + \varepsilon \\X &= X^* + \eta_X \\X^* &= r(Z) + \eta_Z\end{aligned}$$

where the nonparametric function m is the object of interest, X^* is unobservable, Z , X , and Y are observable, $E(\varepsilon|Z, \eta_Z) = E(\eta_X|Z, \eta_Z, \varepsilon) = E(\eta_Z) = 0$, and η_Z and Z are independently distributed.

- Since, in this model,

$$X = r(Z) + \eta_X + \eta_Z$$

and $E(\eta_X + \eta_Z|Z) = 0$, the function r is identified from the joint distribution of (X, Z) .

- The model implies the two moment conditions

$$E(Y|Z = z) = \int m(r(Z) + \eta_Z) dF(\eta_Z)$$

$$E(YX|Z = z) = \int (r(Z) + \eta_Z) m(r(Z) + \eta_Z) dF(\eta_Z)$$

(These moment conditions were used in Newey (2001) to deal with a parametric version of the model with measurement error.)

- Using the representation of these in terms of characteristic functions, Schennach (2007) shows that m and the distribution of X^* are identified.

Shape restrictions on distributions

- Particular shapes or some local conditions on the distributions can often be used to provide identification.
- We provide two examples.

4.3.1. Exchangeability restrictions in the nonadditive model

- Altonji and Matzkin (2005) considered the model

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

where ε is not distributed independently of X , but for some observable variable Z , it is the case that for all x there exists values $z(x)$, $z(x, \bar{x})$ of Z such that for all e

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

- Their leading example is where X denotes the value of a variable for one member of a group, Z denotes the value of the same variable for another member of the same group, and ε , which incorporates the unobservable group effect, is such that its distribution is exchangeable in X and Z , so that for all values t, t' and all e

$$F_{\varepsilon|X=t, Z=t'}(e) = F_{\varepsilon|X=t', Z=t}(e)$$

- In such case, $z(x) = \bar{x}$ and $z(x, \bar{x}) = x$.
- Assume that for all x, z , $F_{\varepsilon|X=x, Z=z}$ is strictly increasing.
- As with the case where ε is assumed to be independent of X , a normalization is needed either on the function m or on the distribution.
- Assume that $m(\bar{x}, \varepsilon) = \varepsilon$.
- Under these assumptions

$$(4.c) \quad m \text{ and } F_{\varepsilon|X=x} \text{ can be recovered from} \\ (F_{Y|X=x, Z=z(x)}, F_{Y|X=\bar{x}, Z=z(x, \bar{x})})$$

- **Proof of (4.c):** Let x and e be given.
- By the strict monotonicity of m in ε ,
 $F_{\varepsilon|X=x,Z=z(x)}(e) = F_{\varepsilon|X=\bar{x},Z=z(x,\bar{x})}(e)$ implies that

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

- Hence, since $m(\bar{x}, e)$, it follows that

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

- Next, since the strict monotonicity of m in ε implies that for all x and e

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

it follows that

$$F_{\varepsilon|X=x}(e) = F_{Y|X=x}(F_{Y|X=x, Z=z(x)}^{-1}(F_{Y|X=\bar{x}, Z=z(x, \bar{x})}(e)))$$



- Rather than imposing a normalization, one may ask what can be identified without imposing any normalization.
- Suppose that the exchangeability condition considered in Altonji and Matzkin (2005) is satisfied.
- Let m, e be given and let $y^* = m(x, e)$.

- Then,

$$m(\bar{x}, e) = F_{Y|X=\bar{x}, Z=x}^{-1} (F_{Y|X=x, Z=\bar{x}}(y^*))$$

and for any x'

$$m(x', e)$$

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

$$= F_{Y|X=x', Z=\bar{x}}^{-1} \left(F_{Y|X=\bar{x}, Z=x'} \left(F_{Y|X=\bar{x}, Z=x}^{-1} (F_{Y|X=x, Z=\bar{x}}(y^*)) \right) \right)$$

- Hence, the effect of changing X from x to x' is

$$m(x', e) - m(x, e) = F_{Y|X=x', Z=\bar{x}}^{-1} \left(F_{Y|X=\bar{x}, Z=x'} \left(F_{Y|X=\bar{x}, Z=x}^{-1} \left(F_{Y|X=x, Z=\bar{x}}(y^*) \right) \right) \right) - y^*$$

4.3.2. Local independence restrictions in the nonadditive model

- Chesher (2003) used a local insensitivity assumption to achieve local identification of the partial derivatives of structural functions in a triangular system of equations.
- To demonstrate a simple version of this restriction, consider a nonadditive model, specified as

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

where m is strictly increasing in ε .

- Suppose that we were interested in inferring the partial derivative of m with respect to X .

- Following arguments analogous to those used in Section 3.3, one can show that for any x, ε

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

- Assuming that all the functions are differentiable, we get that

$$\frac{\partial m^*(\bar{x}, \bar{\varepsilon})}{\partial x} = \left[\frac{\partial F_{Y|X=\bar{x}}(t)}{\partial t} \Big|_{t=m^*(\bar{x}, \bar{\varepsilon})} \right]^{-1} \cdot \left[\frac{\partial F_{Y|X=\bar{x}}(t)}{\partial x} \Big|_{t=m^*(\bar{x}, \bar{\varepsilon})} - \frac{\partial F_{\varepsilon|X=\bar{x}}^*(\bar{\varepsilon})}{\partial x} \right]$$

- The local insensitivity assumption can be stated as the restriction that at $X = \bar{x}$ and $\varepsilon = \bar{\varepsilon}$

$$\frac{\partial F_{\varepsilon|X=\bar{x}}^*(\bar{\varepsilon})}{\partial x} = 0$$

- Assume that the value of $m^*(\bar{x}, \bar{\varepsilon})$ is known.
- It then follows that the derivative of m^* with respect to x , evaluated at $(\bar{x}, \bar{\varepsilon})$, can be identified.

Shape restrictions on functions

- One of the main parts in the specification of an econometric model is the set of restrictions on the functions and distributions of the model.
- We concentrate here on shape restrictions.
- These may prove useful when a specification is such that a particular feature of interest is not identified.
- In such situation, one may consider tightening the set of restrictions by considering particular shapes.
- The analysis of observational equivalence can often be used to determine the search for restrictions that, when added to the model, help to determine identification.
- Economic theory can be used to choose among the possible restrictions.
- We provide some examples.

4.4.1. Homogeneity restrictions

- Homogeneous functions are often encountered in economic models.
- Profit and cost functions of firms in perfectly competitive environments are homogeneous of degree one.
- Production functions are often homogeneous.
- Given the ubiquity of this type of functions, it is worthwhile considering how this restriction can aid in identifying features of a model.
- We provide some examples.

Independent Nonadditive Model

- Consider the Independent Nonadditive Model, described in Section 2.2.1.2, where $Y = m^*(X, \varepsilon)$, m^* is strictly increasing in ε , and ε and X are independently distributed.
- Suppose that we are interested in identifying m^* .
- The analysis of identification in Section 3.3 showed that one can partition the set, Ω , of possible functions m , into classes such that for any two functions, m and \tilde{m} in a class, there exists a strictly increasing $g : R \rightarrow R$ such that for all x, ε

$$\tilde{m}(x, g(\varepsilon)) = m(x, \varepsilon)$$

- Functions within each such class are observationally equivalent, while functions from different classes are not.
- This suggests, then, that any restriction on the set of functions m , which guarantees that for any two different functions in the restricted set, no such g exists, will be sufficient to guarantee identification of m^* within that set.

- Suppose that the function m^* is the profit function of a firm in a perfectly competitive environment, and suppose that (x, ε) is the vector of prices, assumed to possess support R_+^{K+1} .
- Economic theory implies that m^* is continuous and homogenous of degree one in $(x, \varepsilon) \in R_+^{K+1}$.
- Let $(\bar{x}, \bar{\varepsilon})$ denote a specified value of (x, ε) and let $\alpha > 0$ denote a specified number.
- Let Ω denote the set of all functions m that are continuous and homogeneous of degree one and satisfy $m(\bar{x}, \bar{\varepsilon}) = \alpha$.

- Then,

(4.d) if $m, \tilde{m} \in \Omega$ and for some strictly increasing
 $g : R_+ \rightarrow R_+$

$$\tilde{m}(x, g(\varepsilon)) = m(x, \varepsilon)$$

it must be that for all $\varepsilon \in R_+$,

$$g(\varepsilon) = \varepsilon.$$

- **Proof of (4.d)** (Matzkin (2003)): Substituting $x = \bar{x}$ and $\varepsilon = \bar{\varepsilon}$, and using the homogeneity of degree one assumption and the assumption that $\tilde{m}(\bar{x}, \bar{\varepsilon}) = m(\bar{x}, \bar{\varepsilon}) = \alpha$, we get that for all $\lambda > 0$

$$\tilde{m}(\lambda\bar{x}, g(\lambda\bar{\varepsilon})) = m(\lambda\bar{x}, \lambda\bar{\varepsilon}) = \lambda\alpha = \tilde{m}(\lambda\bar{x}, \lambda\bar{\varepsilon})$$

- Since \tilde{m} is strictly increasing in its last coordinate

$$\tilde{m}(\lambda\bar{x}, g(\lambda\bar{\varepsilon})) = \tilde{m}(\lambda\bar{x}, \lambda\bar{\varepsilon}) \quad \text{implies that} \quad g(\lambda\bar{\varepsilon}) = \lambda\bar{\varepsilon}$$

- Since this holds for every $\lambda > 0$, the result follows. ■

- The implication of this result is that in the Independent Nonadditive Model, if we restrict the set to which m^* belongs to be such that all functions, m , in that set are continuous, homogenous of degree one, and satisfy $m(\bar{x}, \bar{\varepsilon}) = \alpha$, then m^* will be identified in that set.

Independent Index Model

- Consider the Independent Index Model, 2.2.1.4, where $Y = m^*(h^*(X), \varepsilon)$, and ε and X are independently distributed.
- The analysis of identification in Section 3.4 showed that one can partition the set, Ω , of possible functions h into classes such that for any two functions, h and \tilde{h} , in a class, there exists a strictly increasing $g : R \rightarrow R$ such that for all x

$$\tilde{h}(x) = g(h(x))$$

- Functions within each such class are observationally equivalent, while functions from different classes are not.
- Hence, any restriction which guarantees that any two function in the restricted set cannot be strictly increasing transformations of each other will suffice to guarantee identification of h^* within that set.

- Let Ω denote the set of all functions $h : X \rightarrow R$ that satisfy the restrictions in the Independent Index Model described in 2.2.1.4 and, in addition, are homogeneous of degree one and satisfy $h(\bar{x}) = \alpha$.
- Assume $h^* \in \Omega$.
- Then,

(4.e) h^* is identified in Ω

- **Proof of (4.e)** (Matzkin (1991b, 1994)): Let $h \in \Omega$.
- Suppose that h is observationally equivalent to h^* .
- Then, by the theorem in Section 3.4, there is some strictly increasing $g : R \rightarrow R$, such that

$$h(x) = g(h^*(x))$$

- Since both $h, h^* \in \Omega$, for all λ

$$\lambda = \left(\frac{\lambda}{\alpha}\right) \alpha = \left(\frac{\lambda}{\alpha}\right) h(\bar{x}) = h\left(\left(\frac{\lambda}{\alpha}\right) \bar{x}\right) = g\left(h^*\left(\left(\frac{\lambda}{\alpha}\right) \bar{x}\right)\right)$$

- The second equality follows by the definition of Ω , the third by the homogeneity of degree one of h , the fourth because for all x , $h(x) = g(h^*(x))$.
- By the homogeneity of degree one of h^* and the specification that $h^*(\bar{x}) = \alpha$, it follows that

$$g\left(\left(\frac{\lambda}{\alpha}\right) h^*(\bar{x})\right) = g\left(\left(\frac{\lambda}{\alpha}\right) \alpha\right) = g(\lambda)$$

- Hence, for all λ , $g(\lambda) = \lambda$.
- Since for all x , $h(x) = g(h^*(x))$, this implies that $h = h^*$.
- Hence, the only function in Ω that is observationally equivalent to h^* is h^* . ■

Discrete Choice Models

- Consider the Discrete Choice Model described in Section 2.2.1.6 with additive unobservables and with the normalization that $V_J(s, z_J, \omega) = 0$.

- Then

$$\Pr(y_j = 0 | s, x_1, \dots, x_J) = F_{\varepsilon_1, \dots, \varepsilon_{J-1}}^* (V_1^*(s, x_1), \dots, V_{J-1}^*(s, x_{J-1}))$$

- From the above analysis it is clear that homogeneity restrictions in each of the V_j^* functions can be used to identify $F_{\varepsilon_1, \dots, \varepsilon_{J-1}}^*$.
- To see this, suppose that the functions V_1^*, \dots, V_{J-1}^* are such that for some \bar{s} , and each j , there exists \bar{x}_j and α_j such that for all s and all λ such that $\lambda \bar{x}_j \in X$, $V_j^*(\bar{s}, \bar{x}_j) = a_j$ and $V_j^*(\bar{s}, \lambda \bar{x}_j) = \lambda a_j$.

- Then, for any (t_1, \dots, t_{J-1}) ,

$$\begin{aligned}
 F_{\varepsilon_1, \dots, \varepsilon_{J-1}}^* (t_1, \dots, t_{J-1}) &= F_{\varepsilon_1, \dots, \varepsilon_{J-1}}^* \left(\left(\frac{t_1}{\alpha_1} \right) \alpha_1, \dots, \left(\frac{t_{J-1}}{\alpha_{J-1}} \right) \alpha_{J-1} \right) \\
 &= F_{\varepsilon_1, \dots, \varepsilon_{J-1}}^* \left(\left(\frac{t_1}{\alpha_1} \right) V_1^* (\bar{s}, \bar{x}_1), \dots, \left(\frac{t_{J-1}}{\alpha_{J-1}} \right) V_{J-1}^* (\bar{s}, \bar{x}_{J-1}) \right) \\
 &= F_{\varepsilon_1, \dots, \varepsilon_{J-1}}^* \left(V_1^* \left(\bar{s}, \left(\frac{t_1}{\alpha_1} \right) \bar{x}_1 \right), \dots, V_{J-1}^* \left(\bar{s}, \left(\frac{t_{J-1}}{\alpha_{J-1}} \right) \bar{x}_{J-1} \right) \right) \\
 &= \Pr \left(y_j = 0 \mid \bar{s}, x_1 = \left(\frac{t_1}{\alpha_1} \right) \bar{x}_1, \dots, x_{J-1} = \left(\frac{t_{J-1}}{\alpha_{J-1}} \right) \bar{x}_{J-1} \right)
 \end{aligned}$$

- Hence, $F_{\varepsilon_1, \dots, \varepsilon_{J-1}}^*(t_1, \dots, t_{J-1})$ can be recovered from $\Pr\left(y_J = 0 \mid \bar{s}, x_1 = \left(\frac{t_1}{\alpha_1}\right) \bar{x}_1, \dots, x_{J-1} = \left(\frac{t_{J-1}}{\alpha_{J-1}}\right) \bar{x}_{J-1}\right)$ as long as this conditional probability is identified.
- When $F_{\varepsilon_1, \dots, \varepsilon_{J-1}}^*$ is identified, one can recover each V_g^* function as in Matzkin (1991a).
- (See Matzkin and Newey (1993) and Lewbel and Linton (2007) for the use of homogeneity restrictions when $J = 2$.)

4.4.2. Additivity Restrictions

- As with homogeneous functions, additive functions also appear often in economic models.
- Aggregate demand is the sum of individual demands; cost functions are sums of fixed cost and variable cost functions; total income is the sum of income from work and income from other sources.
- We describe below two particular examples where additivity can be used to identify nonparametric functions.

4.4.2.1. Additivity in conditional expectations

- Consider an additive model, where for unknown functions m_1^* and m_2^* ,

$$E(Y|X = (x_1, x_2)) = m_1^*(x_1) + m_2^*(x_2)$$

- Following the arguments in Linton and Nielsen (1995), one can show that

(4.f) m_1^* and m_2^* can be recovered, up to at an additive constant, from $E(Y|X = (x_1, x_2))$

- **Proof of (4.f):** Note that

$$\begin{aligned}\int E(Y|X = (x_1, x_2)) f(x_2) dx_2 &= \int (m_1^*(x_1) + m_2^*(x_2)) f(x_2) dx_2 \\ &= m_1^*(x_1) + \int m_2^*(x_2) f(x_2) dx_2\end{aligned}$$

- Hence, once one specifies a value for $\int m_2^*(x_2) f(x_2) dx_2$, one can obtain $m_1^*(x_1)$ for all x_1 .
- For each x_2 , the value of $m_2^*(x_2)$ can then be obtained by

$$\begin{aligned}m_2^*(x_2) &= E(Y|X = (x_1, x_2)) - m_1^*(x_1) \\ &= E(Y|X = (x_1, x_2)) \\ &\quad - \int E(Y|X = (x_1, x_2)) f(x_2) dx_2 \\ &\quad + \int m_2^*(x_2) f(x_2) dx_2\end{aligned}$$

which depends on the same constant $\int m_2^*(x_2) f(x_2) dx_2$. ■

4.4.2.2. Additivity in a known function

- When a nonparametric function can only be identified up to a strictly increasing transformation, a scale as well as a location normalization will be necessary.
- An often convenient way of imposing these is to assume that the nonparametric function is linearly additive in one of the coordinates, the coefficient of that coordinate is known, and the value of the subfunction of the other coordinates is specified at one point.
- In other words, partition X into subvectors X_1, \dots, X_J , so that $X_1 \in R$, and $X = (X_1, \dots, X_J) \in R^K$.

- Suppose that for functions h_2^*, \dots, h_J^* ,

$$h^*(X) = X_1 + \sum_{j=2}^J h_j^*(X_j)$$

and that for some value $(\bar{x}_2, \dots, \bar{x}_J)$ of (X_2, \dots, X_J) , the value of $\sum_{j=2}^J h_j^*(\bar{x}_j)$ is specified, then,

(4.g) if h^*, \bar{h} are two functions satisfying these restrictions,

h^, \bar{h} cannot be strictly increasing transformations of each other*

- **Proof of (4.g):** Let $g:R \rightarrow R$ be a strictly increasing function.
- Suppose that for all X , $h^*(X) = g(\tilde{h}(X))$.
- Then, letting $X = (x_1, \bar{x}_2, \dots, \bar{x}_J)$, it follows that for all x_1 ,
$$g\left(x_1 + \sum_{j=2}^J \tilde{h}_j(\bar{x}_j)\right) = x_1 + \sum_{j=2}^J h_j^*(\bar{x}_j).$$
- Since $\sum_{j=2}^J \tilde{h}_j(\bar{x}_j) = \sum_{j=2}^J h_j^*(\bar{x}_j)$, it follows that g must be the identity function. ■
- This result can be used in the nonadditive model, the nonadditive index model, and discrete choice models, using arguments similar to the ones used for the homogeneity of degree one case.

Restrictions on functions and distributions

- Often, a combination of restrictions on functions and distributions is used.
- We provide some examples below.

4.5.1. Control functions

- A control function is a function of observable variables such that conditioning on its value purges any statistical dependence that may exist between the observable and unobservable explanatory variables in an original model.
- The *control function* approach was fully developed, and analyzed for parametric selection models, in Heckman and Robb (1985).
- The method is commonly used for identification of models where the explanatory observable variables, X , and the explanatory unobserved variables, ε , are not independently distributed.
- In this method, the unobservable, ε , is modeled as a function of observed or identified variables, W , which have independent variation from the endogenous explanatory variables, X .
- We provide an example.

4.5.1.1. A control function in an additive model

- Newey, Powell and Vella (1999) considered identification and estimation of the model

$$Y = m(X, Z_1) + \varepsilon$$

with the additional equation

$$X = \pi(Z) + u$$

and the restrictions

$$E[\varepsilon|u, Z] = E[\varepsilon|u] \quad \text{and} \quad E[u|Z] = 0$$

where Z_1 is a subvector of Z .

- (See also Ng and Pinkse (1995) and Pinkse (2000).)

- Since, in this model, $E[\varepsilon|u] = E[\varepsilon|u, Z] = E[\varepsilon|u, X, Z]$, u can be used as a control function to identify m .
- Since $E[u|Z] = 0$, the function π can be recovered from the joint distribution of (X, Z) .
- Hence, $u = X - \pi(Z)$ can also be recovered.
- Moreover, the structure of the model implies that for some g

$$\begin{aligned} E[Y|X, Z] &= m(X, Z_1) + E[\varepsilon|u] \\ &= m(X, Z_1) + g(X - \pi(Z)) \end{aligned}$$

- The following identification result is established in Newey, Powell and Vella (1999):
- **Theorem 4.5** (Newey, Powell and Vella (1999)): *Suppose that $m(x, z_1)$, $g(u)$, and $\pi(Z)$ are differentiable, the boundary of the support of (Z, u) has zero probability, and with probability one, $\text{rank}(\partial\pi(Z_1, Z_2)/\partial Z_2) = d_X$, where d_X denotes the dimension of d_X .*
- *Then, $m(X, Z_1)$ is identified (up to constant).*

- As noted in Newey, Powell, and Vella (1999), one can use the additive structure to derive the derivatives of the functions m directly.
- Let $h(X, Z_1, Z_2) = E[Y|X, Z_1, Z_2]$.

- Then, since

$$h(X, Z_1, Z_2) = m(X, Z_1) + g(X - \pi(Z))$$

it follows that

$$\frac{\partial h(X, Z_1, Z_2)}{\partial X} = \frac{\partial m(X, Z_1)}{\partial X} + \frac{\partial g(u)}{\partial u} \Big|_{u=X-\pi(Z)}$$

$$\begin{aligned} \frac{\partial h(X, Z_1, Z_2)}{\partial Z_1} &= \frac{\partial m(X, Z_1)}{\partial Z_1} \\ &\quad - \left(\frac{\partial \pi(Z_1, Z_2)}{\partial Z_1} \right)' \frac{\partial g(u)}{\partial u} \Big|_{u=X-\pi(Z)} \end{aligned}$$

$$\frac{\partial h(X, Z_1, Z_2)}{\partial Z_2} = - \left(\frac{\partial \pi(Z_1, Z_2)}{\partial Z_2} \right)' \frac{\partial g(u)}{\partial u} \Big|_{u=X-\pi(Z)}$$

- Assume that $\text{rank}(\partial\pi(Z_1, Z_2)/\partial Z_2) = d_X$.
- Define

$$D(Z) = \left[\left(\frac{\partial\pi(Z_1, Z_2)}{\partial Z_2} \right) \left(\frac{\partial\pi(Z_1, Z_2)}{\partial Z_2} \right)' \right]^{-1} \left(\frac{\partial\pi(Z_1, Z_2)}{\partial Z_2} \right)$$

- Then, multiplying $\partial h(X, Z_1, Z_2) / \partial Z_2$ by $D(Z)$ and solving gives

$$\frac{\partial m(X, Z_1)}{\partial X} = \frac{\partial h(X, Z_1, Z_2)}{\partial X} - D(Z) \frac{\partial h(X, Z_1, Z_2)}{\partial Z_2}$$

$$\begin{aligned} \frac{\partial m(X, Z_1)}{\partial Z_1} &= \frac{\partial h(X, Z_1, Z_2)}{\partial Z_1} \\ &\quad + \left(\frac{\partial \pi(Z_1, Z_2)}{\partial Z_1} \right)' D(Z) \frac{\partial h(X, Z_1, Z_2)}{\partial Z_2} \end{aligned}$$

- The above gives identification of mup to an additive constant.

- An additional restriction is necessary to identify such a constant.
- Suppose, for example, that $E[\varepsilon] = 0$.
- Then, as shown in Newey, Powell, and Vella (1999), for any function $\tau(u)$ such that $\int \tau(u) du = 1$,

$$\begin{aligned} & \int E[Y|X, Z_1, u] \tau(u) du - E \left[\int E[Y|X, Z_1, u] \tau(u) du \right] \\ &= m(X, Z_1) - E[m(X, Z_1)] + E[Y] \\ &= m(X, Z_1) \end{aligned}$$

- Hence, the constant of m is identified.

4.5.2. Linear factor models

- When the unobservable vector ε in a model is driven by factors that are common to some equations, one might want to use a factor model.
- Factor models were introduced into economics by Jöreskog and Goldberger (1972), Goldberger (1972), Chamberlain and Griliches (1975), and Chamberlain (1977a,b).
- (See Aigner, Hsiao, Kapteyn, and Wansbeek (1984) for an in-depth review and analysis.)
- The standard situation analyzed in factor models is the one where there are L measurements on K mutually independent factors arrayed in a vector θ .

- Let G denote the vector of measurements.
- Then, the model is specified as

$$G = \mu + \Lambda\theta + \delta$$

where G is $L \times 1$, θ is independent of δ , μ is an $L \times 1$ vector of means, which may depend on a vector of observable variables X , θ is $K \times 1$, δ is $L \times 1$, and Λ is $L \times K$, the coordinates of $\delta = (\delta_1, \dots, \delta_L)$ are assumed to be mutually independent, as well as the coordinates of $\theta = (\theta_1, \dots, \theta_K)$, and δ and θ are assumed to be independent.

- Anderson and Rubin (1956) discuss the identification problem in factor models.
- More recently, Carneiro, Hansen, and Heckman (2003) have shown that factor models can be identified when the matrix Λ has a particular structure.
- Bonhomme and Robin (2006) analyze identification using the third and fourth moments of the distributions of the measurements.

- Carneiro, Hansen, and Heckman (2003) consider a system of L measurements on K factors,

$$M_1 = m_1(X) + \beta_{11}\theta_1 + \cdots + \beta_{1K}\theta_K + \delta_1$$

$$M_2 = m_2(X) + \beta_{21}\theta_1 + \cdots + \beta_{2K}\theta_K + \delta_2$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$M_L = m_L(X) + \beta_{L1}\theta_1 + \cdots + \beta_{LK}\theta_K + \delta_L$$

where $\delta = (\delta_1, \dots, \delta_L)$, $E(\delta) = 0$, and where $\theta = (\theta_1, \dots, \theta_K)$ is distributed independently of δ .

- A special case that they consider is one where there are two or more measurements devoted exclusively to factor θ_1 , and at least three measurements that are generated by factor θ_1 , two or more further measurements that are devoted only to factors θ_1 and θ_2 , with at least three measurements that depend on θ_1 and θ_2 , and so fourth, in blocks of at least two.
- Order G under this assumption so that

$$\Lambda = \begin{pmatrix} 1 & 0 & \cdot & 0 & \dots & 0 \\ \lambda_{21} & 0 & \cdot & 0 & \dots & 0 \\ \lambda_{31} & 1 & 0 & 0 & \dots & 0 \\ \lambda_{41} & \lambda_{42} & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & 0 & \dots & 0 \\ \lambda_{L1} & \lambda_{L2} & \lambda_{L3} & \dots & \dots & \lambda_{LK} \end{pmatrix}$$

- Assuming nonzero covariances,

$$\text{Cov}(g_j, g_l) = \lambda_{j1} \lambda_{l1} \sigma_{\theta_1}^2 \quad l = 1, 2; j = 1, \dots, L; j \neq l$$

where $G = (g_1, \dots, g_L)$.

- In particular,

$$\begin{aligned}\text{Cov}(g_1, g_l) &= \lambda_{l1} \sigma_{\theta_1}^2 \\ \text{Cov}(g_2, g_l) &= \lambda_{l1} \lambda_{21} \sigma_{\theta_1}^2\end{aligned}$$

- Hence, assuming that $\lambda_{l1} \neq 0$, one obtains

$$\lambda_{21} = \frac{\text{Cov}(g_2, g_l)}{\text{Cov}(g_1, g_l)}$$

- It follows that from $\text{Cov}(g_1, g_l) = \lambda_{21}\sigma_{\theta_1}^2$, one can obtain $\sigma_{\theta_1}^2$, and hence λ_{l1} , $l = 1, \dots, L$.
- One can then proceed to the next set of two measurements and identify

$$\text{Cov}(g_l, g_j) = \lambda_{l1}\lambda_{j1}\sigma_{\theta_1}^2 + \lambda_{l2}\lambda_{j2}\sigma_{\theta_2}^2 \quad l = 3, 4; j \geq 3; j \neq l$$

- Since we can know the first term on the right hand side by the previous arguments, we can proceed using $Cov(g_l, g_j) - \lambda_{l1}\lambda_{j1}\sigma_{\theta_1}^2$ and identify the λ_{j2} $j = 1, \dots, L$, using similar arguments.
- Proceeding in this fashion, one can identify Λ and the variance of θ , Σ_{θ} , subject to diagonal normalizations.
- Knowing Λ and Σ_{θ} , one can identify the variance, D_{δ} , of δ .
- Next, using the mutual independence of the factors θ_i ($i = 1, \dots, K$), one can identify the densities of each θ_i .

- To provide a simple case, developed in Carneiro, Hansen, and Heckman (2003), suppose that

$$G_1 = \lambda_{11}\theta_1 + \delta_1$$

$$G_2 = \lambda_{21}\theta_1 + \delta_2$$

where $\lambda_{11} = 1$ and $\lambda_{21} \neq 0$.

- Subject to the normalization that $\lambda_{11} = 1$, λ_{21} is identified.
- Thus, one can write these equations as

$$\begin{aligned}G_1 &= \theta_1 + \delta_1 \\ \frac{G_2}{\lambda_{21}} &= \theta_1 + \left(\frac{\delta_2}{\lambda_{21}} \right)\end{aligned}$$

where θ_1 , δ_1 , and (δ_2/λ_{21}) are mutually independent.

- By Kotlarski (1967), one can nonparametrically identify the densities of θ_1 , δ_1 , and (δ_2/λ_{21}) .
- The next equations in the system

$$G_3 = \lambda_{31}\theta_1 + \theta_2 + \delta_3$$

$$G_4 = \lambda_{41}\theta_1 + \lambda_{42}\theta_2 + \delta_4$$

can be written as

$$G_3 - \lambda_{31}\theta_1 = \theta_2 + \delta_3$$

$$\frac{G_4 - \lambda_{41}\theta_1}{\lambda_{42}} = \theta_2 + \left(\frac{\delta_4}{\lambda_{42}}\right)$$

where θ_2 , δ_3 , and (δ_4/λ_{42}) are mutually independent.

- Again, one can apply Kotlarski's theorem.
- Proceeding in this fashion, all the densities are identified.
- From the knowledge about the densities of θ_i and the factor loadings, one can apply standard deconvolution methods to nonparametrically identify the δ terms in the model.
- Cunha, Heckman, and Matzkin (2004) extend this analysis to factor models of the type

$$Y_t = m_t(X, \beta_t\theta + \delta_t) \quad t = 1, \dots, T$$

where m_t is strictly increasing in its last argument.

- Assuming that $(\theta, \delta_1, \dots, \delta_T)$ is distributed independently of X and that at some specified value \bar{x}_t of X ,

$$m_t(\bar{x}_t, \beta_t\theta + \delta_t) = \beta_t\theta + \delta_t$$

one can recover the distribution of $\eta_t = \beta_t\theta + \varepsilon_t$, and the function m_t , since, by previous arguments

$$F_{\eta_t}(\eta_t) = F_{Y_t|X_t=\bar{x}_t}(\eta_t)$$

and

$$m_t(x_t, \eta_t) = F_{Y_t|X_t=x_t}^{-1}(F_{Y_t|X_t=\bar{x}_t}(\eta_t))$$

- Let r_t denote the inverse of m_t with respect to η_t .
- Then, given y_t, x_t

$$\eta_t = r_t(x_t, y_t) = F_{Y_t|X_t=\bar{x}_t}^{-1}(F_{Y_t|X_t=x_t}(y_t))$$

- We can then analyze the identification of the factor model, as in Carneiro, Hansen, and Heckman (2003), from the system

$$\eta_t = \beta_t \theta + \varepsilon_t$$

where η_t is interpreted as a measurement on θ .

- One could also allow X to depend on η_t , using Matzkin (2004).
- Suppose that there exists Z_t such that η_t is independent of X_t conditional on Z_t .
- Then, one can obtain identification of m_t and η_t .
- One way of guaranteeing that this condition is satisfied is by assuming that there exists an unobservable ϕ_t and a function v_t , such that

$$X_t = v_t(Z_t, \phi_t)$$

and ϕ_t is independent of (θ, δ_t) conditional on Z_t .

4.5.3. Index Models with fixed effects

- Abrevaya (2000) established the identification of the coefficients of a linear index model for panel data models with two observations.
- Abrevaya's model was

$$Y_{it} = D \circ G(\beta X_{it}, \varepsilon_i, \eta_{it}) \quad i = 1, \dots, N; \quad t = 1, 2$$

where for each ε_i , the function G is strictly increasing in βX_{it} and η_{it} .

- The function D is assumed to be monotone increasing and nonconstant, (η_{i1}, η_{i2}) is independent of $(X_{i1}, X_{i2}, \varepsilon_i)$ and has support R^2 , and one of the coordinates of $X_{it} \in R^K$ is continuously distributed with support R , conditional on the other coordinates.

- The model is then like the one studied in Han (1987) with the added fixed effect ε_i .
- In the same way that Matzkin (1991b) modified the arguments in Han (1987) to show the identification of a nonparametric index function, one can modify Abrevaya's arguments to establish the identification of the nonparametric function h^* in the model

$$Y_{it} = D \circ G(h^*(X_{it}), \varepsilon_i, \eta_{it}) \quad i = 1, \dots, N; \quad t = 1, 2$$

- Assume that the function G is strictly increasing in its first and third arguments; the function D is monotone increasing and nonconstant; (η_{i1}, η_{i2}) is independent of $(X_{i1}, X_{i2}, \varepsilon_i)$; conditional on ε_i , (X_{i1}, η_{i1}) is independent of (X_{i2}, η_{i2}) ; and (X_{i1}, X_{i2}) has support R^{2K} .
- Let h^* belong to a set of continuous, homogeneous of degree one functions, $h : R^K \rightarrow R$, that are strictly increasing in the last coordinate, and satisfy $h(\bar{x}) = \alpha$.
- Then, within this set,

(4.h) h^* is identified.

- **Proof of (4.h):** Suppose that h belongs to the set of continuous, homogeneous of degree one functions, that are strictly increasing in the last coordinate, and satisfy $h(\bar{x}) = \alpha$, and that $h \neq h^*$.
- Then, following the arguments in Matzkin (1991b), one can show that there exist neighborhoods N_1 and N_2 such that for all $x_1'' \in N_1$ and $x_2'' \in N_2$,

$$h^*(x_1'') > h^*(x_2'') \quad \text{and} \quad h(x_1'') < h(x_2'')$$

- For each ε_i , the model is as the one considered in Matzkin (1991b).

- Hence, by analogous arguments, it follows by independence that, conditional on ε_i , since $h^*(x_1'') > h^*(x_2'')$

$$\begin{aligned} & \Pr [Y_{it} > Y_{is} | X_{it} = x_1'', X_{is} = x_2'', \varepsilon_i; h^*] \\ = & \Pr \{ (\eta_{it}, \eta_{is}) | D \circ G (h^*(x_1''), \varepsilon_i, \eta_{it}) > D \circ G (h^*(x_2''), \varepsilon_i, \eta_{is}) \} \\ > & \Pr \{ (\eta_{it}, \eta_{is}) | D \circ G (h^*(x_1''), \varepsilon_i, \eta_{it}) < D \circ G (h^*(x_2''), \varepsilon_i, \eta_{is}) \} \\ = & \Pr [Y_{it} < Y_{is} | X_{it} = x_1'', X_{is} = x_2'', \varepsilon_i; h^*] \end{aligned}$$

- And, since $h(x_1'') < h(x_2'')$,

$$\begin{aligned} & \Pr [Y_{it} > Y_{is} | X_{it} = x_1'', X_{is} = x_2'', \varepsilon_i; h] \\ &= \Pr \{ (\eta_{it}, \eta_{is}) | D \circ G (h(x_1''), \varepsilon_i, \eta_{it}) > D \circ G (h(x_2''), \varepsilon_i, \eta_{is}) \} \\ &< \Pr \{ (\eta_{it}, \eta_{is}) | D \circ G (h(x_1''), \varepsilon_i, \eta_{it}) < D \circ G (h(x_2''), \varepsilon_i, \eta_{is}) \} \\ &= \Pr [Y_{it} < Y_{is} | X_{it} = x_1'', X_{is} = x_2'', \varepsilon_i; h] \end{aligned}$$

- Integrating over any two possible distributions for ε_i conditional on (x_1'', x_2'') , we get

$$\Pr[Y_{it} > Y_{is} | X_{it} = x_1'', X_{is} = x_2''; h^*] >$$

$$\Pr[Y_{it} < Y_{is} | X_{it} = x_1'', X_{is} = x_2''; h^*]$$

and

$$\Pr[Y_{it} > Y_{is} | X_{it} = x_1'', X_{is} = x_2''; h] <$$

$$\Pr[Y_{it} < Y_{is} | X_{it} = x_1'', X_{is} = x_2''; h]$$

- Hence, the distribution of the observable variables is different under h than under h^* .
- It follows that h^* is identified. ■
- Chesher (2005) considers a model with many unobservables.

4.5.4. Single equation models with multivariate unobservables

- Matzkin (2003) considers the model

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

where $(\varepsilon_1, \dots, \varepsilon_K)$ is independent of X and $\varepsilon_1, \dots, \varepsilon_K$ are mutually independent.

- Suppose that X can be partitioned into (X_1, \dots, X_K) such that for some known r , and unknown functions m_1, \dots, m_K

$$Y = r(m_1(X_1, \varepsilon_1), m_2(X_2, \varepsilon_2), \dots, m_K(X_K, \varepsilon_K))$$

- Suppose that r is strictly increasing in each coordinate and that for each k , there exist for all coordinate j different from k , values $x_j^{(k)}$ such that, when

$x = \left(x_1^{(k)}, \dots, x_{k-1}^{(k)}, x_k, x_{k+1}^{(k)}, \dots, x_K^{(k)} \right)$ the conditional distribution $F_{Y|X=x}$ of Y given $X = \left(x_1^{(k)}, \dots, x_K^{(k)} \right)$ is strictly increasing and identified, and for all $j \neq k$,

$$m_j \left(x_j^{(k)}, \varepsilon_j \right) = \alpha_j$$

for a specified value α_j .

- Then, for all x_k and ε_k

$$\begin{aligned} F_{Y|X} &= \left(x_1^{(k)}, \dots, x_{k-1}^{(k)}, x_k, x_{k+1}^{(k)}, \dots, x_K^{(k)} \right) \left(r \left(\alpha_1, \dots, \alpha_{k-1}, m_k \left(x_k, \varepsilon_k \right), \alpha_{k+1}, \dots, \alpha_K \right) \right) \\ &= F_{\varepsilon_k} \left(\varepsilon_k \right) \end{aligned}$$

- In this expression, all functions and values are known except for $m_k(x_k, \varepsilon_k)$ and $F_{\varepsilon_k}(\varepsilon_k)$.
- A normalization on either of these, as described in Section 3.3, or a restriction on m_k , as described in Section 4.1.1, can be used to identify m_k and F_{ε_k} .
- A similar argument can be used to show that under analogous conditions, all the functions m_k and all the marginal distributions F_{ε_k} can be identified.
- Since $\varepsilon_1, \dots, \varepsilon_K$ are assumed to be mutually independent, the identification of the marginal distributions of each of the ε_k implies the identification of $F_{\varepsilon_1, \dots, \varepsilon_K}$.

- To provide an example, suppose that

$$Y = \sum_{k=1}^K m_k(x_k, \varepsilon_k)$$

where for each k , all ε_k , and for specified values $\alpha_1, \dots, \alpha_K$, \tilde{x}_k and \bar{x}_k , $m_k(\tilde{x}_k, \varepsilon_k) = \alpha_k$ and $m_k(\bar{x}_k, \varepsilon_k) = \varepsilon_k$.

- Then, letting $x^* = (\tilde{x}_1^{(k)}, \dots, \tilde{x}_{k-1}^{(k)}, \bar{x}_k, \tilde{x}_{k+1}^{(k)}, \dots, \tilde{x}_K^{(k)})$,
 $x^{**} = (\tilde{x}_1^{(k)}, \dots, \tilde{x}_{k-1}^{(k)}, x_k, \tilde{x}_{k+1}^{(k)}, \dots, \tilde{x}_K^{(k)})$

$$m_k(x_k, \varepsilon_k) = F_{Y|X=x^{**}}^{-1} \left(F_{Y|X=x^*}(\varepsilon_k + \sum_{j=1; j \neq k}^K \alpha_j) \right) - \sum_{j=1; j \neq k}^K \alpha_j$$

- Note that the linear random coefficients model, where $Y = \sum_{k=1}^K \beta_k x_k$, for unobservable, mutually independent β_1, \dots, β_K , is an example of a model that satisfies the above restrictions.
- In this case, $\tilde{x}_k = 0$ and $\bar{x}_k = 1$.

Conclusions

- This chapter has attempted to provide some insight into some of the results that have been developed recently for nonparametric models, with emphasis on those with nonadditive unobservable random terms.
- We first presented some general identification results about nonparametric models with additive unobservables, nonadditive unobservables, index models, simultaneous equations models, and discrete choice models.
- Next, we discussed some techniques that have been used to achieve identification, such as imposing additional restrictions on the functions and/or distributions in the models, or augmenting the data.