Readings on Demand Identification

Manu Navjeevan

August 29, 2020

Contents

Starting with some readings on nonparametric identification of simultaneous equations for Demand Identification in Two Sided Markets

1 Matzkin Identification Chapter

Here mainly focus on section 3.5, Identification in Simultaneous Equation Models. Based off of Matzkin (2008), which should also be covered in these notes. *Main Results here are Theorems 1 and 2 from Matzin (2008).*

Focus is on the simultaneous equations model, where $Y \in \mathbb{R}^G$ denotes a vector of observable dependent variables, $X \in \mathbb{R}^K$ denotes a vector of observable explanatory variables, $\epsilon \in \mathbb{R}^G$ denotes a vector of unobservable explanatory variables and the relationship between these vectors is specified by a function $r^* : \mathbb{R}^G \times \mathbb{R}^K \to \mathbb{R}^G$ such that

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

The set S of $r^*, F_{\epsilon,X}$ that are considered consist of twice differentiable functions $r : \mathbb{R}^G \times \mathbb{R}^K \to \mathbb{R}^G$ and twice differentiable, strictly increasing distributions $F_{\epsilon,X} : \mathbb{R}^G \times \mathbb{R}^K \to \mathbb{R}$ such that (i) for all $F_{\epsilon,X}, \epsilon$ and X are distributed independently of each other (ii) for all r and 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 $\epsilon = r(y,x)$, and (iv) for all r, all $F_{\epsilon,X}$ and all x, the distribution of Y given $X = x$, induced by r and $F_{\epsilon|X=x}$ has support $\mathbb{R}^G$.

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

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

This is the reduced form system of the structural equations system determined by r. Will let h^* denote the reduced form function determined by r^* (the "true" value of 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,

$$\epsilon = AY + BX$$

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

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

where $\Pi = -A^{-1}B$ and $\nu = A^{-1}\epsilon$. The identification of the true values A^*, B^* is well studied (Koopmans (1949), Koopmans, Rubin, Leipnik (1950), and Fisher (1966) as well as most econometrics textbooks).

- My guess is that full nonparametric identification would amount to the conditions for identification of the linear system holding locally, evverywhere.

Main results here, assume that $\mathbb{E}(\epsilon) = 0$ and $\text{Var}(\epsilon) = \Sigma^*$, an unkown matrix. Let W denote the varianve of ν. The identification of (A^*, B^*, Σ^*) is achieved when it can be uniqely recovered from Π and $\text{Var}(\nu)$. A priori restrictions on A^*, B^*, Σ^* are typically used to determine the existence of a unique solution for any element of the above triple.

Analagoulsy, one can obstain necessary and suffecient 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 from the definition of observational equivalence, can state that two fucntions $r, \tilde{r}$ satisfying (i)-(iv) are obs. equivalent iff $\exists f_\epsilon, \tilde{f}_\epsilon$ such that $(f_\epsilon, r), (\tilde{f}_\epsilon, \tilde{r}) \in S$ and for all y, x

$$\tilde{f}_\epsilon(\tilde{r}(y,x)) \left| \frac{\partial \tilde{r}(y,x)}{\partial y} \right| = f_\epsilon(r(y,x)) \left| \frac{\partial r(y,x)}{\partial y} \right| \tag{1}$$

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

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

where h is the reduced form equation corresponding to r above. Since

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

it follows from assumption (ii) that $\left| \frac{\partial g(\epsilon, x)}{\partial \epsilon} \right| > 0$ everywhere. 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 into (1), we can see that $(\tilde{r}, \tilde{f}_\epsilon) \in S$ is observationally equivalent to $(r, f_\epsilon) \in S$ iff $\forall \epsilon, x$

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

The following theorem provides conditions garunteeing a transformation g of ϵ does not generate an observationally equivalent pair $(\tilde{r}, \tilde{f}_\epsilon)$.

Theorem 1 (Matzkin, 2005)**.** *Let $(r, f_\epsilon) \in S$. Let $g(\epsilon, x)$ be such that $\tilde{r}(y, x) = g(r(y, x), x)$ and $\tilde{\epsilon} = g(\epsilon, x)$ are such that $(\tilde{r}, \tilde{f}_\epsilon) \in S$. If for some ϵ, x, the rank of the matrix*

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

Alternatively, can express this as an identification result for the function r^*

Theorem 2 (Matzkin, 2005)**.** *Let $M \times \Gamma$ denote the set of pairs $(r, f_\epsilon) \in S$. The function r^* is identified in M if $r^* \in M$ and, for all $f_\epsilon \in \Gamma$ and all $\tilde{r}, r \in M$ such that $\tilde{r} \neq r$, there exist y, x such that the matrix*

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

is strictly larger than G, 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 1. As a simple example, consider the simultaneous equations model analyzed by Matzkin (2007c), where for some unknown function g^* and some parameter values β^*, γ^*,

$$y_1 = g^*(y_2) + \epsilon_1$$
$$y_2 = \beta^* y_1 + \gamma^* x + \epsilon_2$$

Further, assume that (ϵ_1, ϵ_2) has an everywehre positive, differentiable desnsity $f^*_{\epsilon_1, \epsilon_2}$ such that, for two not

necessarily known a-priori values $(\bar{\epsilon}_1, \bar{\epsilon}_2)$ and $(\epsilon_1'', \epsilon_2'')$

$$0 \neq \frac{\partial \log f_{\epsilon_1, \epsilon_2}^*(\bar{\epsilon}_1, \bar{\epsilon}_2)}{\partial \epsilon_1} \neq \frac{\partial \log f_{\epsilon_1, \epsilon_2}^*(\epsilon_1'', \epsilon_2'')}{\partial \epsilon_1} \neq 0$$

$$\frac{\partial \log f_{\epsilon_1, \epsilon_2}^*(\bar{\epsilon}_1, \bar{\epsilon}_2)}{\partial \epsilon_2} = \frac{\partial \log f_{\epsilon_1, \epsilon_2}^*(\epsilon_1'', \epsilon_2'')}{\partial \epsilon_2} = 0$$

The observable exogeneous variable x is assumed to be distributed independently of (ϵ_1, ϵ_2) and to possess support $\mathbb{R}$. In this model

$$\epsilon_1 = r_1^*(y_1, y_2, x) = y_1 - g^*(y_2)$$
$$\epsilon_2 = r_2^*(y_1, y_2, x) = -\beta^* y_1 + y_2 - \gamma^* x$$

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 so long as $1 > \beta^* \partial g^*(y_2)/\partial y_2$. Since the first element in the diagonal is positive, it follows from Gale and Nikaido (1965) that the function r^* is globally invertible if the condition $1 > \beta^* \frac{\partial g^*(y_2)}{\partial y_2}$ holds for all y^1. Let $r, \tilde{r}$ be any two differentiable functions satisfying this condition and the other properties assumed about r^*. Suppose that at some y_2, $\frac{\partial \tilde{g}(y_2)}{\partial y_2} \neq \frac{\partial g(y_2)}{\partial y_2}$. Assume also that $\gamma \neq 0$ and $\tilde{\gamma} \neq 0$. Let $f_{\epsilon_1, \epsilon_2}$ denote any density satisfying the same properties that $f_{\epsilon_1, \epsilon_2}^*$ is assumed to satisfy. Denote by (ϵ_1, ϵ_2) and $(\epsilon_1', \epsilon_2')$ the two points such that

$$0 \neq \frac{\partial f_{\epsilon_1, \epsilon_2}(\epsilon_1, \epsilon_2)}{\partial \epsilon_1} \neq \frac{\partial \log f_{\epsilon_1, \epsilon_2}(\epsilon_1, \epsilon_2')}{\partial \epsilon_1} \neq 0$$

$$\frac{\partial \log f_{\epsilon_1, \epsilon_2}(\epsilon_1, \epsilon_2)}{\partial \epsilon_2} = \frac{\partial \log f_{\epsilon_1, \epsilon_2}(\epsilon_1', \epsilon_2')}{\partial \epsilon_2} = 0$$

Define

$$a_1(y_1, y_2, x) := \frac{\partial \log f_{\epsilon_1, \epsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \epsilon_1} - \beta \frac{\partial \log f_{\epsilon_1, \epsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \epsilon_2}$$

$$a_2(y_1, y_2, x) := \left(\frac{\frac{\partial^2 g(y_2)}{\partial y_2^2}}{1 - \beta \frac{\partial g(y_2)}{\partial y_2}} - \frac{\frac{\partial^2 \tilde{g}(y_2)}{\partial y_2^2}}{1 - \beta \frac{\partial \tilde{g}(y_2)}{\partial y_2}} \right) - \frac{\partial g(y_2)}{\partial y_2} \frac{\partial \log f_{\epsilon_1, \epsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \epsilon_1}$$

$$+ \frac{\partial \log f_{\epsilon_1, \epsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \epsilon_2}$$

$$a_3(y_1, y_2, x) := -\gamma \frac{\partial \log f_{\epsilon_1, \epsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \epsilon_2}$$

By Theorem 3.4, r and $\tilde{r}$ will not be observationally equivalent if for all $f_{\epsilon_1, \epsilon_2}$ there exists (y_1, x) such that the rank of the matrix

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

---

[1]like a global extension to the implicit function theorem

is 3. Let

$$a_1'(y_1, y_2, x) := \left(\tilde{\beta} - \beta\right) \frac{\partial \log f_{\epsilon_1, \epsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \epsilon_2}$$

$$a_2'(y_1, y_2, x) := \left(\frac{\frac{\partial^2 g(y_2)}{\partial y_2^2}}{1 - \beta \frac{\partial g(y_2)}{\partial y_2}} - \frac{\frac{\partial^2 \tilde{g}(y_2)}{\partial y_2^2}}{1 - \beta \frac{\partial \tilde{g}(y_2)}{\partial y_2}}\right) + \left(\frac{\partial \tilde{g}(y_2)}{\partial y_2} - \frac{\partial g(y_2)}{\partial y_2}\right) \left(\frac{\partial \log f_{\epsilon_1, \epsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \epsilon_1}\right)$$

$$a_3'(y_1, y_2, x) = (\tilde{\gamma} - \gamma) \frac{\partial \log f_{\epsilon_1, \epsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \epsilon_2}$$

Multiplying the first column of A by $-\frac{\partial \log f_{\epsilon_1, \epsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \epsilon_1}$ and adding it to the third column, and multiplying the second column by $\frac{\partial \log f_{\epsilon_1, \epsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \epsilon_2}$ and adding it to the third column[2], one obtains the matrix

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

By assumption either

$$a_2'(\bar{y}_1, \bar{y}_2, \bar{x}) \neq 0 \quad \text{or} \quad a_2'(\tilde{y}_1, \tilde{y}_2, \tilde{x}) \neq 0$$

where $(\bar{y}_1, \bar{y}_2, \bar{x})$ correspond to an arbitrary (ϵ_1, ϵ_2), and $(\tilde{y}_1, \tilde{y}_2, \tilde{x})$ correspond to $(\epsilon_1'', \epsilon_2'')$ from above.[3]

Suppose the later. Let $y_1 = g(y_2 + \epsilon_1)$ and let $x = \frac{-\beta y_1 + y_2 - \epsilon_2'}{\gamma}$. Follows that

$$\frac{\partial \log f_{\epsilon_1, \epsilon_2}(y_1 - g(y_2), -\beta y_1 + y_2 - \gamma x)}{\partial \epsilon_2} = 0$$

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

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

so the derivatives of g^* are identified.

---

[2]Therse are standard row operations and do not change the invertibility
[3]Actually I'm a bit unsure here

2   An Almost Ideal Demand System

Here I read the 1980 Paper, *An Almost Ideal Demand System* By Angus Deaton and John Muelbauer. The paper goes over how to identify demand systems in large markets. The hope is to adapt this somehow to two sided markets under competition.

2.1   Introduction

- Richard Stone (1954) first estimated a system of demand equations derived explicitly from consumer theory

- Paper proposes and estimates a new model which is of comparable generality to the Rotterdam and translog models but has advantages over both

- Model (AIDS) gives an arbitrarty first order-approximation to any demand system, satisfies the axioms of choice exactly, aggregates perfectly over consumers without invoking parallel linear Engle curves, has a functional form which is consistent with known household-budget data, simple to estimate, largely avoids need for non-linear estimation.

- Model is estimated on postwar British data.

2.2   Specification of the AIDS

Generally, starting point has been the specification of a function general enough to act as a second-order approximation to any arbitrart direct or indirect utility function. It is possible to use a first order approximation to the demand functions themselves as in the Rotterdam model.

AIDS approach follows from these approaches but builds froma specific class of preferences by which the theorems of Muellbauer permit exact aggregation over consumers. Preferences are known as PIGLOG preferences and are represented via expenditure function[1]. Denote the expenditure function $c(u,p)$ for utility u and price vector p and define the PIGLOG class by

$$\log c(u,p) = (1-u)\log a(p) + u\log b(p); \quad \text{for} \ \ u \in [0,1] \tag{1}$$

The positive constant returns to scale functions $a(p)$ and $b(p)$ can be regarded as the costs of substinence and bliss, respectively. Next take specific functional forms for $\log a(p)$ and $\log b(p)$. To ensure a flexible functional form for the cost function, the specific funcitonal form for the above functions must posess enough parameters so that at any single point the cost function derivatives up to a second order w.r.t u and p can be set equal to those of an arbitrary cost function[2]. Take

$$\log a(p) = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_k \sum_j \gamma_{kj}^* \log p_k \log p_j \tag{2}$$

$$\log b(p) = \log a(p) + \beta_0 \prod_k p_k^{\beta_k} \tag{3}$$

With these specifications, the AIDS cost function is written

$$\log c(u,p) = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_k \sum_j \gamma_{kj}^* \log p_k \log p_j + u\beta_0 \prod_k p_k^{\beta_k} \tag{4}$$

where $\alpha_i, \beta_i, \gamma_{ij}^*$ are the parameters of the model. Clearly, $c(u,p)$ is CRS in p provided that $\sum_i \alpha_i = 1, \sum_j \gamma_{kj}^* = \sum_k \gamma^* kj = \sum_j \beta_j = 0$. Also straightforward to check that (4) has enough paramaters to be a flexible functional form remembering that utility is orginal so we can always choose a normalization such that, at a point $\frac{\partial^2 \log c}{\partial u^2} = 0$.

---

[1] minimum expenditure necessary to obtain a specific utility level at given prices

[2] Approximation is at a point, basically means that the model specification gives the "specifier" full control over the first and second derviatives *at one specific point*

Demand equations can be derived from the cost functions in equation (4). It is a fundamental property of the cost function that its price derivatives are the quantities demanded $\frac{\partial c(u,p)}{\partial p_i} = q_i$. Multiplying both sides by $p_i/c(u,p)$ delivers

$$\frac{\partial \log c(u,p)}{\partial p_i} = \frac{p_i q_i}{c(u,p)} = w_i \tag{5}$$

where w_i is the budget share of good i. So log differentiation of (4) gives budget shares as a function of prices and utility

$$w_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + u\beta_0\beta_i \prod_k p_k^{\beta_k} \tag{6}$$

where $\gamma_{ij} = \frac{1}{2}(\gamma_{ij}^* + \gamma_{ji}^*)^3$. For a utility mazimizing consumer, total expenditure x is equal to $c(u,p)$ and this equality can be inverted to five u as a function of p and x, the indirect utility function. Do this for (4) and substitute into (6) to obtain budget shares as a function of prices (p) and total expenditure (x); these are the AIDS demand functions in budget share form.

$$w_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log(c/P) \tag{7}$$

where P is a price index defined by

$$\log P = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2}\sum_{j,k} \gamma_{kj} \log p_k \log p_j \tag{8}$$

The restrictions on the parameters of (4) plus the definition of γ_{ij} imply restrictions on the parameters of the AIDS equation (7). Specifically, they require

$$\sum_i \alpha_i = 1, \quad \sum_i \gamma_{ij} = 0, \quad \sum_i \beta_i = 0 \tag{9}$$

$$\sum_j \gamma_{ij} = 0 \tag{10}$$

$$\gamma_{ij} = \gamma_{ji} \tag{11}$$

provided (9), (10), (11) hold, (7) represents a system of demand functions which add up to total expenditure $(\sum w_i = 1)$, are homogeneous of degree 0 and which satisfy Sltusky symmetry.

Given these, the AIDS is simply interpreted: in the absence of changes in relative prices and the real expenditure (x/P), the budget shares are constants. This is a starting point for predictions using the model[4].

Changes in relative prices work through the terms γ_{ij}: each γ_{ij} represents 10^2 the effect on the i-th budget share of a 1 percent increse in the j-th price with (x/P) held constant. Changes in real expenditure operate through the β_i coefficients; these add to zero and are positive for "luxuries" and negative for "necessities".

2.3 Aggregation Over Households

Note: This section basically shows that the AIDS demand system derived above for a single household aggregates nicely under some assumptions, and that the aggregate demand curve looks basically like the individual household demand curve. In general, I don't think that microfounding the aggregate demand model in this way is so important, espcially given the strict assumption required to do this, but it does provide some intuition as to underlying assumptions that may be useful.

Aggregation theory from Muellbauer (1975, 1976) imply that exact aggregation is possible if, for each house-

---

[3]If the Slutsky matix is symmetric $\gamma_{ji}^* = \gamma_{ij}^* = \gamma_{ij} = \gamma_{ji}$ anyways.

[4]This seems like a testable resttiction of the model

hold h, behavior is descrbed by a generalization of (7) given below[1].

$$w_{ih} = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log(x_h/k_h P) \tag{12}$$

k_h can be interpreted as a sophisticated measure of households size, which, in principle, could account for age composition, other household characteristics, and economies of household size. This allows for a limited amount of taste variation across households.

The share of aggregate expenditure on good i in the aggregate budget of all households, denoted $\bar{\omega}_i$ is given by

$$\sum_h p_i q_{ih} \Big/ \sum x_h = \sum_h x_j w_{ih} \Big/ \sum x_h$$

Applied to (12) this yields

$$\bar{w}_i = \alpha_i + \sum_j \gamma_{ij} \log p_j - \beta_i \log P + \beta_i \left\{ \sum_h x_j \log(x_h/k_h) \Big/ \sum x_h \right\} \tag{13}$$

Define the aggregate index k by

$$\log(\bar{x}/k) := \sum_h x_h \log(x_h/k_h) \Big/ \sum x_h \tag{14}$$

where $\bar{x}$ is the average level of total expenditure x_h. So (13) becomes

$$\bar{w}_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log(\bar{x}/kP) \tag{15}$$

Notice this is the same form as in (12) and shows that under these assumptions, aggregate budget shares correspond to the decisions of a representative household whose budget is given by $\bar{x}/k$, the "representative" budget level.

When estimating the model, it is generally assumed that k is constant or uncorrelated with $\bar{x}$ or p so that no omitted variably bias occurs from omitting k from estimation and redefining $\alpha_i^* = \alpha_i - \beta_i \log k^*$, where k^* is the constant or sample mean value of k.

2.4   Generality of the Model

Flexible functional form property of the AIDS cost function implies that the demand functions derived from it are first order approximations to any set of demand functions derived from utility maximizing. If mazimizing behavior is not assumed but demands are continuous functions of the budget and prices, then AIDS demand functions (7) can still be viewerd as a first-order approximation[1]. Because AIDS model still gives full control over first and second derivatives, this is not bad.

However, there are still a lot of parameters. One obvious restriction is that, for some pairs $(i,j), \gamma_{ij}$ should be zero. For such pairs, the budget share of each is independent of the price of the other if (x/P) is held constant. Can be shown that γ_{ij} has approximately the same sign as the compensated cross-price elasticity[2].

---

[1]Interesting to note which parameters depend on h here

[1]I think this is the best interpretation of the AIDS model.

[2]This paper is published in 1980, well before the 1996 Tibshirani LASSO paper. Restricting some of the γ_{ij} to be 0 is essentially a sparsity condition. It would be interesting to come up with a way of applying LASSO here. Maybe also take a Bayseian approach. The researcher has some prior on which products are 0, use LASSo to update the prior.

2.5   Restrictions

Starting from equations (7) and (8) as maintained hypotheses can examine the effects of restrictions (9)-(11). The conditions (9) are the adding-up restrictions and, as can be checked from (7), they ensure that $\sum w_i = 1$.

Homogeneity of the demand functions require restriction (10) which can be tested equation by equation. Symmetry is also a testable restriction.

2.6   Estimation

In general, estimation can be carried out by substituting (8) into (7) to give

$$w_i = (\alpha_i - \beta_i\alpha_0) + \sum_j \gamma_{ij}\log p_j + \beta_i\left\{\log x - \sum_k \alpha_k\log p_k - \frac{1}{2}\sum_{k,j}\gamma_{kj}\log p_k\log p_j\right\} \tag{16}$$

and estimating this non-linear system of equations by maximum likelihood or other methods with and without the restrictions (10) and (11). With rnough data, this is not particularly difficult to estimato since the first order conditions for MLE are linear in α and γ, given β, and vice versa so that concentration allows iteration on a subset of the parameters.

Although all the parameters in (16) are identified given sufficient variation in the independent variables, in many examples the practical identification of α_0 may be difficult. The parameter is only identified from the α_i's in (16) by the presence of these latter inside the term in braces, originally in the formula for $\log P$. However, in situations where individual prices are closely collinear, $\log P$ is unlikely to be very sensitive to its weights so that changes in the intercept term in (16) due to variations in α_0 can be offset in the α's with minimal effect on $\log P$. This can be overcome in practice by assigning a value to α_0 a priori.

In many situations it is possobile to explot the collinearity of prices for a much simplet estimation technique. Note from (7) that if P is known, the model is linear in it's parameters and so estimation can be done equation by equation through OLS (at least without cross-equation restrictions such as symmetry.)

The adding-up constraints (10) will be automatically satisfied by these estimates. In situations where prices are closely collinear, it may be adequate to approximate P as proportional to some index P^* such as $\log P^* = \sum w_k\log p_k$. In this case, (7) can be estimated as

$$w_i = (\alpha_i - \beta_i\log\phi) + \sum_j \gamma_{ij}\log p_j + \beta_i\log\left(\frac{x}{P^*}\right) \tag{17}$$

In this setup, the α_i parameters are only identifies up to a scalar multiple of β_i[1]. If we write $\alpha_i^* = \alpha_i - \beta_i\log\phi$, it is easily seen that $\sum\alpha_k^* = 0$ is required for adding up, since $\sum\beta_k = 0$.

Empirical results show that (17) is a good approximation for (16). However, it is still an approximation.

2.7   An Application to Postwar British Data

Estimate the model using annual British daa from 1954 to 1974 on eight nondurable groups of consumer expenditure, namely: food, clothing, housing services, fuel, drink + tobacco, transport + communication services, other goods, and other services.

As discussed above, if we assume that the index k in (15) is either constant or that its deviations are independently distributied from those of the average budget $\bar{x}$ and of prices, no biases result from its omission. In particular, allow the intercepts in (15) to absorb the $-\beta_i\log k$ terms. Then proceed by first following the strategy outlined in Section 2.6, setting $\log P^* = \sum w_k\log p_k$ for each year and estimation equation (17).

---

[1]because we don't know what ϕ is presumably

3   Semiparametric Instrumental Variable Estimation of Treatment Response Models *Alberto Abadie (JoE, 2003)*

3.1   Introduction

Economists long concrened with how to estimate the effect of a treatment on some outcome of interest, possilbe after conditioning on a vector of covariates. Main empirical challenge in studies of this type arises from the fact that selection for treatment is usually related to the potential outcomes that individuals would attain with and without the treatment.

Variety of methods have been proposed to overcome the selection problem. Traditional approach relies on distributional assumptions and functional form restrictions to identify average treatment effects and other treatment parameters of interest. Estimators based on this approach can be seriously biased by modest departures from parametric assumptions. In addition, a number of researchers have noted that strong parametric assumptions are not necessary to identify parameters of interest.

Consequently, desirable to develop robust estimators of treatment parameters bases on nonparametric or semi parametric identification procedures. Article introduces a new class of IV estimators of linear and nonlinear average treatment response models with covariates. In the spirit of Roehrig (1988), identification is attained non-parametrically and does not depend on the choice of parametric model. As in the IV model of IA (1994), identification comes from a binary instrument that induces exogenous selection into treatment for some subset of the population. Approach taken here easily accommodates covariates and can be used to estimate nonlinear models with a binary and endogenous regressor.

Ability to control for covariates is important because instruments may require conditioning on a set of covariates to be valid. Covariates can also be used to reflect observable differences in the composition of populations. As a by-product of the general framework introduced here, develop an IV estimator that provides a linear least squares approximation to an average treatment response function, just as OLS provides a linear least squares approximation to a conditional expectation. Shown that 2SLS typically does not have this property.

3.2   The Framework

3.2.1   Identification Problem

Suppose interested in the effect of some treatment, represented by the binary variable D (college) on some outcome of interest Y (earnings). As in Rubin (1974, 1977) define Y_1, Y_0 as the potential outcomes that an individual would attain with and without being exposed to the treatment. Treatment parameters are defiend as characteristics of the distribution of (Y_1, Y_0) for well defined subpopulations.

In the example, Y_1 represents potential earnings as a college graduate while Y_0 represents potential earnings as a non-graduate. Treatment effect is $Y_1 - Y_0$. Now, an identification problem arises from the fact that we cannot observe both potential outcomes Y_1 and Y_0 for the same individual, only observe $Y = Y_1 D + Y_0 (1 - D)$. Since one of the potential outcomes is always missing, cannot compute the treatment effect for any individual.

However, comparisons of average earnings of average effect of the treatment on the treated and non treated to not usually give the right answer

$$\begin{aligned}
\mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0] &= \mathbb{E}[Y_1|D=1] - \mathbb{E}[Y_0|D=0] \\
&= \mathbb{E}[Y_1 - Y_0|D=1] + \big(\mathbb{E}[Y_0|D=1] - \mathbb{E}[Y_0|D=0]\big)
\end{aligned} \tag{1}$$

Because treatment status is not independent of potential outcomes, this cannot be reduced to ATE. The first term of the RHS of (1) gives the average effect of the treatment on the treated. Second term represents bias caused by selection into treatment.

Of course, can think of other parameters of interest.

3.2.2 Identification by Instrumental Variables

IV methods proposed to recover treatment parameters. Article follows the approach of Imbens and Angrist (1994).

Suppose that there is a binary instrument Z available to the researcher. The formal requisites to be an instrument are stated below. Informally, the role of an instrument is to induce and exogenous variation in the treatment variable. The binary variable D_z represents potential treatment status given $Z = z$. Suppose, for example, that Z is an indicator for college proximity. Then $D_0 = 0$ and $D_1 = 1$, living nearby a college at the end of high school, but would not graduate otherwise. The treatment status indicator variable can then be expressed as $D = ZD_1 + (1 - Z)D_0$. In practice, observe Z and D, but not the potential treatment indicators. Following terminology of Angrist (1996), population is divided into groups by treatment indicators, compliers, never takers, always takers, defiers.

In order to state the properties that a valid instrument should have, need to include Z in the definition of the potential outcomes. For a particular individual, the variable Y_{zd} represents the potential outcome would obtain if $D_0 = 0$ for some individual.

Assumption 1. Assume the following on the joint distribution of observables and unobservables
1. Independence of the instrument: Conditional on X, the random vector $(Y_{00}, Y_{01}, Y_{10}, D_0, D_1)$ is independent of the instrument Z.
2. Exclusion of the instrument: $\mathbb{P}(Y_{1d} = Y_{0d}|X) = 1$ for $d \in \{0, 1\}$
3. First Stage: $0 < \mathbb{P}(Z = 1|X) < 1$ and $\mathbb{P}(D_1 = 1|X) > \mathbb{P}(D_0 = 1|X)$
4. Monotonicity: $\mathbb{P}(D_1 \geq D_0|X) = 1$

Assumptions are essentially the conditional versions of those used in Angrist. Assumption 2.1.2 means that we can write potential outcomes in terms of D.

Imbens and Angrist (1994) show that if assumption 2.1 holds, in absence of covariates, a simple IV estimand identifies the LATE

$$\alpha_{IV} = \frac{\text{cov}(Y, Z)}{\text{cov}(D, Z)} = \frac{\mathbb{E}[Y|Z = 1] - \mathbb{E}[Y|Z = 0]}{\mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]} = \mathbb{E}[Y_1 - Y_0|D_1 > D_0] \tag{2}$$

3.3 Identification of Statistical Characteristics for Compliers

Section presents an identification theorem that includes previous results on IV models for treatment effects on IV models for treatment effects as special cases. To study identification, proceed as if we know the joint distribution of (Y, D, X, Z). In practice, can use a random sample from (Y, D, X, Z) to construct empirical analogs.

Lemma 1. *Under Assumption 1*

$$\mathbb{P}(D_1 > D_0|X) = \mathbb{E}[D|Z = 1, X] - \mathbb{E}[D|Z = 0, X] > 0$$

Proof. Under Assumption 1

$$
\begin{aligned}
\mathbb{P}(D_1 > D_0|X) &= 1 - \mathbb{P}(D_1 = D_0 = 1|X) - \mathbb{P}(D_1 = D_0 = 1|X) && \text{(by monotonicity)} \\
&= 1 - \mathbb{P}(D_1 = D_0 = 0|X, Z = 1) - \mathbb{P}(D_1 = D_0 = 1|X, Z = 0) && \text{(by independence of } Z) \\
&= 1 - \mathbb{P}(D = 0|Z, Z = 1) - \mathbb{P}(D = 1|X, Z = 0) && \text{(by monotonicity)} \\
&= \mathbb{P}(D = 1|X, Z = 1) - \mathbb{P}(D = 1|X, Z = 0) && \text{(because } D \text{ is binary)} \\
&= \mathbb{E}[D|X, Z = 1] - \mathbb{E}[D|X, Z = 0] && \text{(because } D \text{ is binary)}
\end{aligned}
$$

$\square$

Lemma says that, under assumption 2.1, proportion of compliers in the population is identified given X and that this proportion is greater than 0. Preliminary result is important for establishing the following theorem.

Theorem 1. *Let $g(\cdot)$ be any measurable function of (Y, D, X) such that $\mathbb{E}[|g(Y, D, X)|] < \infty$. Define*

$$\kappa_{(0)} \equiv (1 - D)\frac{(1 - Z) - \mathbb{E}[(1 - Z)|X]}{\mathbb{E}[(1 - Z)|X]\mathbb{E}[Z|X]}$$

$$\kappa_{(1)} \equiv D\frac{Z - \mathbb{E}[Z|X]}{\mathbb{E}[(1 - Z)|X]\mathbb{E}[Z|X]}$$

$$\kappa \equiv \kappa_{(0)}\mathbb{E}[(1 - Z)|X] + \kappa_{(1)}\mathbb{E}[Z|X]$$

$$= 1 - \frac{D(1 - Z)}{\mathbb{P}(Z = 0|X)} - \frac{(1 - D)Z}{\mathbb{P}(Z = 1|X)}$$

Then under Assumption 1

$$\mathbb{E}[g(Y, D, X)|D_1 > D_0] = \frac{1}{\mathbb{P}(D_1 > D_0)}\mathbb{E}\left[\kappa g(Y, D, X)\right] \tag{3}$$

$$\mathbb{E}[g(Y_0, X)|D_1 > D_0] = \frac{1}{\mathbb{P}(D_1 > D_0)}\mathbb{E}\left[\kappa_{(0)}g(Y, X)\right] \tag{4}$$

$$\mathbb{E}[g(Y_1, X)|D_1 > D_0] = \frac{1}{\mathbb{P}(D_1 > D_0)}\mathbb{E}\left[\kappa_{(1)}g(Y, X)\right] \tag{5}$$

This theorem is a powerful identification result, says that any statistical characteristic that can be defined in terms of moments of the joint distribution is identified for compliers. *Since D is exogenous given X for compliers, Theorem 1 can be used to identify meaningful treatment parameters for used to identify meaningful treatment parameters for this group of the population.*

Now go through the proof of this Theorem.

Proof. Monotonicity along with the law of total expectation implies that

$$\mathbb{E}[g(Y, D, X)|X, D_1 > D_0] = \frac{1}{\mathbb{P}(D_1 > D_0|X)}\Big\{\mathbb{E}[g(Y, D, X)|X]$$
$$- \mathbb{E}[g(Y, D, X)|X, D_1 = D_0 = 1]\mathbb{P}(D_1 = D_0 = 1|X)$$
$$- \mathbb{E}[g(Y, D, X)|X, D_1 = D_0 = 0]\mathbb{P}(D_1 = D_0 = 0|X)\Big\}$$

Since Z is ignorable and independent of the potential outcomes given X and since monotonicity is assumed, the above equation can be rewritten as

$$\mathbb{E}[g(Y, D, X)|X, D_1 > D_0] = \frac{1}{\mathbb{P}(D_1 > D_0|X)}\Big\{\mathbb{E}[g(Y, D, X)|X]$$
$$- \mathbb{E}[g(Y, D, X)|X, D = 1, Z = 0]\mathbb{P}(D = 1|X, Z = 0)$$
$$- \mathbb{E}[g(Y, D, X)|X, D = 0, Z = 1]\mathbb{P}(D = 0|X, Z = 1)\Big\}$$

Consider similarly, by law of total expectation

$$\mathbb{E}[D(1 - Z)g(Y, D, X)|X] = \mathbb{E}[g(Y, D, X)|X, D = 1, Z = 0]\mathbb{P}(D = 0, Z = 1|X)$$
$$= \mathbb{E}[g(Y, D, X)|X, D = 1, Z = 0]\mathbb{P}(D = 1|X, Z = 0)\mathbb{P}(Z = 0|X)$$
$$\implies \frac{1}{\mathbb{P}(Z = 0|X)}\mathbb{E}[D(1 - Z)g(Y, D, X)|X] = \mathbb{E}[g(Y, D, X)|X, D = 1, Z = 0]\mathbb{P}(D = 1|X, Z = 0)$$

$$\mathbb{E}[Z(1 - D)g(y, D, X)|X] = \mathbb{E}[g(Y, D, X)|X, D = 0, Z = 1]\mathbb{P}(D = 0, Z = 1)$$
$$= \mathbb{E}[g(Y, D, X)|X, D = 0, Z = 1]\mathbb{P}(D = 0|X, Z = 1)\mathbb{P}(Z = 1|X)$$
$$\implies \frac{1}{\mathbb{P}(Z = 1|X)}\mathbb{E}[Z(1 - D)g(y, D, X)|X] = \mathbb{E}[g(Y, D, X)|X, D = 0, Z = 1]\mathbb{P}(D = 0|X, Z = 1)$$

Under Assumption 1.3 can combine the last three equations to

$$\mathbb{E}[g(Y,D,X)|X,D_1>D_0] = \frac{1}{\mathbb{P}(D_1>D_0|X)}\mathbb{E}\left[g(Y,D,X)\left(1-\frac{D(1-Z)}{\mathbb{P}(Z=0|X)}-\frac{Z(1-D)}{\mathbb{P}(Z=1|X)}\right)\Big|X\right]$$

Bayes' Theorem implies that

$$\mathbb{P}(D_1>D_0|X) = \frac{f_{X|D_1>D_0}(x)\mathbb{P}(D_1>D_0)}{f_X(x)}$$

$$\Longrightarrow \frac{1}{\mathbb{P}(D_1>D_0|X)} = \frac{f_X(x)}{f_{X|D_1>D_0}(x)\mathbb{P}(D_1>D_0)}$$

Applying Bayes' Theorem and Integrating yields

$$\int \mathbb{E}[g(Y,D,X)|X,D_1>D_0]dP(X|D_1>D_0)$$

$$= \int \mathbb{E}[g(Y,D,X)|X,D_1>D_0]f_{X|D_1>D_0}(x)dx$$

$$= \frac{1}{\mathbb{P}(D_1>D_0)}\int \mathbb{E}\left[g(Y,D,X)\left(1-\frac{D(1-Z)}{\mathbb{P}(Z=0|X)}-\frac{Z(1-D)}{\mathbb{P}(Z=1|X)}\right)\Big|X\right]dP(X)$$

or

$$\mathbb{E}[g(Y,D,X)|D_1>D_0] = \frac{1}{\mathbb{P}(D_1>D_0)}\mathbb{E}[\kappa g(Y,D,X)]$$

This proves the first part of the theorem. To prove the second part, note that

$$\begin{aligned}
\mathbb{E}[g(Y,X)(1-D)|X,D_1>D_0] &= \mathbb{E}[g(Y_0,X)|D=0,X,D_1>D_0]\mathbb{P}(D=0|X,D_1>D_0) \\
&= \mathbb{E}[g(Y_0,X)|Z=0,X,D_1>D_0]\mathbb{P}(D=0|X,D_1>D_0) \quad \text{(for compliers } Z=D) \\
&= \mathbb{E}[g(Y_0,X)|X,D_1>D_0]\mathbb{P}(Z=0|X) \qquad\qquad \text{(by independence of } Z)
\end{aligned}$$

The proof of the next parts follows quickly from here. For the second equality note that

$$\begin{aligned}
\mathbb{E}[g(Y_0,X)|X,D_1>D_0] &= \mathbb{E}\left[g(Y,X)\frac{(1-D)}{\mathbb{P}(Z=0|X)}\Big|X,D_1>D_0\right] \\
&= \frac{1}{\mathbb{P}(D_1>D_0|X)}\mathbb{E}[\kappa\frac{(1-D)}{\mathbb{P}(Z=0|X)}g(Y,X)|X] \\
&= \frac{1}{\mathbb{P}(D_1>D_0|X)}\mathbb{E}[\kappa_0 g(Y,X)|X]
\end{aligned}$$

Integration of this equation as before yields the result. □

3.4   Estimation of Average Response Functions

3.4.1   Local Average Response Functions

Consider the function of (D,X) that is equal to $\mathbb{E}[Y_0|X,D_1>D_0]$ if $D=0$ and is equal to $\mathbb{E}[Y_1|X,D_1>D_0]$ if $D=1$. This function describes average treatment responses for any group of compliers defined by some value for the covariates. Refer to this function as the local average response function (LARF). Since $Z=D$ for compliers, under Assumptions 1.1 and 1.2, Z is ignorable for the compliers given X. It follows that

$$\mathbb{E}[Y|X,D=0,D_1>D_0] = \mathbb{E}[Y_0|D_1>D_0]$$

$$\mathbb{E}[Y|X,D=1,D_1>D_0] = \mathbb{E}[Y_1|X,D_1>D_0]$$

$$\mathbb{E}[Y|X,D=1,D_1>D_0] - \mathbb{E}[Y|X,D=0,D_1>D_0] = \mathbb{E}[Y_1-Y_0|X,D_1>D_0]$$

So $\mathbb{E}[Y|X, D_1 > D_0]$ is the LARF. Important special case arises when $\mathbb{P}(D_0 = 0|X) = 1$. This happens in some randomized experiments.

The face that the conditional expectation of Y given D and X holds for compliers has an interpretation as an average treatment response function would not be very useful in the absence of Theorem 1.

3.4.2 Estimation

Section describes two ways to learn about the LARF: (i) estimate a parameterization of the LARF by Least Squares, (ii) Specify a parametric distribution for $P(Y|X, D, D_1 > D_0)$ and estimate the parameters of the LARF by a Maximum Likelihood. Since identification does not depend on these parametric specifications have an appealing interpretations under misspecification of the parameterization in (i) or (ii). Through this $W = (Y, D, X, Z)$ and $\{w_i\}_{i=1}^n$ is a sample of realizations of W.

Least Squares Suppose that the LARF belongs to some class of parametric functions $\mathcal{H} = \{h(D, X; \theta) : \theta \in \Theta \subset \mathbb{R}^m\}$ in the Lebesgue space of square-integrable functions. Let θ_0 be the vector of parameters such that $\mathbb{E}[Y|X, D, D_1 > D_0] = h(D, X; \theta)$. Then

$$\begin{aligned}
\theta_0 &= \arg\min_{\theta \in \Theta} \mathbb{E}\left[(Y - h(D, X; \theta))^2 | D_1 > D_0 \right] \\
&= \arg\min_{\theta \in \Theta} \mathbb{E}\left[\kappa (Y - h(D, X; \theta))^2 \right]
\end{aligned} \tag{6}$$

Under functional form misspecification (LARF does not belong to $\mathcal{H}$), θ_0 are the parameters of the best least squares approximation from $\mathcal{H}$ to $\mathbb{E}[Y|D, X, D_1 > D_0]$:

$$\theta_0 = \arg\min_{\theta \in \Theta} \mathbb{E}\left[\left(\mathbb{E}[Y|D, X, D_1 > D_0] - h(D, X; \theta) \right)^2 \right]$$