Identification of Models of the Labor Market

Eric French and Christopher Taber,

Federal Reserve Bank of Chicago and Wisconsin

November 6, 2009

What is this about?

Our original plan was a chapter on "treatment effects"

Upon further thought, we were not sure this made much sense

There have been many survey papers done on treatment effects already

Do we really need another paper going through ATE,TT,LATE,MTE,TUT etc.?

Our goal was to stick to the basic theme but move in a different direction

- We start with identification of the "treatment effect model" or as it is sometimes called the "generalized Roy model"
- However, rather than go from there to think about all of the different treatment effects
- we instead will use the ideas to study different models

Basically we will think about non-paramateric identification of labor market models (but not much about non-parametric estimation)

Is this interesting enough for a handbook chapter?

- Speaking for myself, I think so. I always begin a research project by thinking about nonparametric identification.
- Literature on nonparametric identification not particularly highly cited-particularly by labor economists
- At the same time this literature has had a huge impact on the field. A Heckman two step model without an exclusion restriction is often viewed as highly problematic these days-presumably because of nonparametric identification

Basic Strategy

Our general goal will be to start with the Roy model (the labor supply version of it where you only get to observe the wage in one sector)

Slowly go through the identification arguments in that case

Then think about how this applies in other models.

12 Measurement Error

13 Peer Effects

Definition of Identification

We follow Matzkin's (2007) formal definition of identification and follow her notation exactly (in the text we try to explain this so that it is comprehensible to a second year PhD student).

Define

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

$\psi^* \in \Omega$ *is identified in the model* S *if for any* $\psi \in \Omega$,

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

The Roy Model

Here we follow Heckman and Honore (1990) closely

Basic model is workers can hunt or fish

They simply choose the occupation with higher income.

A worker chooses to fish if $Y_{fi} > Y_{ri}$ and thus the workers observed income is

$$Y_i = \max\{Y_{fi}, Y_{ri}\}.$$

The econometrician generally observes Y_i and job choice $(J_i = f$ or $r)$

Heckman and Honore show that the model is identified under normality assumptions

but the most general nonparametric model is not identified

We focus on the version of the model

$$\begin{aligned} Y_{fi} &= g_f(X_{fi}, X_{0i}) + \varepsilon_{fi} \\ Y_{ri} &= g_r(X_{ri}, X_{0i}) + \varepsilon_{ri}. \end{aligned}$$

where $(\varepsilon_{fi}, \varepsilon_{ri})$ is independent of (X_{fi}, X_{ri}, X_{0i})

Level of Generality

Another issue in a chapter such as this is the level of generality to use.

One level is

$$Y_{fi} = X_i' \beta_f + \varepsilon_{fi}$$

where u_i is normally distributed.

Another is

$$Y_i = g(X_i, \varepsilon_{fi}).$$

We focus throughout the chapter on the intermediate case

$$Y_{fi} = g_f(X_i) + \varepsilon_{fi}$$

We focus on the “labor supply” version of this in which you only observe the wage in one sector (Y_{fi})

In this case fishing is like market work and hunting is like home production

We will first formally state the result, but then (slowly) go through the intuition

Assumption

$(\varepsilon_{fi}, \varepsilon_{ri})$ is continuously distributed with distribution function G, support $\mathbb{R}^2$, and is independent of (X_{0i}, X_{ri}, X_{fi}).

Assumption

The support of $(g_f(X_{fi}, x_{0i}), g_r(X_{ri}, x_{0i}))$ is $\mathbb{R}^2$ for all x_0 in the support of X_0.

Assumption

The marginal distributions of ε_{fi} and ε_{ri} have medians equal to zero.

Theorem

If $(J_i \in \{f, r\}, Y_{fi}$ *if* $J_i = f, X_{fi}, X_{ri}, X_{0i})$ *are all observed, under assumptions 2.1,2.2, and 2.3,* g_f*,* g_r*, and* G *are identified on their support.*

Identification takes 4 steps

1. Identification of reduced form choice model
2. Identification of the wage equation g_f
3. Identification of structural choice model g_r
4. Identification of joint distribution of error terms

These four basic stages are used in identification of most models. We will draw on this for the rest of the chapter.

Step 1: Identification of Choice Model

This part is well known in a number of papers (Manski and Matzkin being the main contributors) We can write the model as

$$
\begin{aligned}
Pr(J_i = f \mid X_{i0}, X_{if}, X_{ir}) &= Pr(\varepsilon_{ir} - \varepsilon_{if} < g_f(X_{if}, X_{i0}) - g_r(X_{ir}, X_{i0})) \\
&= G_{r-f}(g^*(X_{if}, X_{ir}, X_{i0})),
\end{aligned}
$$

where G_{r-f} is the distribution function for $\varepsilon_{ir} - \varepsilon_{if}$ and $g^*(X_{if}, X_{ir}, X_{i0}) \equiv g_f(X_{if}, X_{i0}) - g_r(X_{ir}, X_{i0})$.

Given data only on choices, the model is only identified up to a montonic transformation. Let M be any strictly increasing function, then

$$g^{*}(X_f, X_r, X_0) > \varepsilon_r - \varepsilon_f$$

if and only if

$$M(g^{*}(X_f, X_r, X_0)) > M(\varepsilon_r - \varepsilon_f).$$

A very convenient normalization is to choose the uniform distribution for $\varepsilon_r - \varepsilon_f$. Then

$$\begin{aligned} Pr(J_i = f \mid X_f, X_r, X_0) &= Pr(M(\varepsilon_r - \varepsilon_f) < \widehat{g}^*(X_f, X_r, X_0)) \\ &= \widehat{g}^*(X_f, X_r, X_0). \end{aligned}$$

Thus we have thus established that we can write the model as $J_i = f$ if and only if $\widehat{g}^*(X_{if}, X_{ir}, X_{i0}) > \widehat{\varepsilon}_i$ where $\widehat{\varepsilon}_i$ is uniform $(0, 1)$ and that $\widehat{g}^*$ is identified.

Step 2: Identification of the Wage Equation g_f

Next consider identification of $g_{f.}$ This is basically the standard selection problem.

Notice that we can identify the distribution of Y_f conditional on $(X_f, X_r, X_0, J_i = f.)$

In particular we can identify

$$Med(Y_i \mid X_{fi}, X_{ri}, X_{0i}, J_i = f) = g_f(X_{fi}, X_{0i}) + Med(\varepsilon_{fi} \mid X_{fi}, X_{ri}, X_{0i}, \widehat{\varepsilon}_i < \widehat{g}^*(X_{fi}, X_{ri}, X_{0i})).$$

Exclusion restriction is key, we need a variable X_{ri} that allows us move (X_{fi}, X_{0i}) holding $\widehat{g}^*$ and thus $Med(\varepsilon_{fi} \mid X_{fi}, X_{ri}, X_{0i}, \widehat{\varepsilon}_i < \widehat{g}^*(X_{fi}, X_{ri}, X_{0i}))$ fixed.

Identification at Infinity

What about the location?

Notice that

$$\begin{aligned}
&\lim_{\widehat{g}^*(x_f,x_r,x_0)\to 1} Med(Y_f \mid X_f = x_f, X_r = x_r, X_0 = x_0, J = f) \\
&= g_f(x_f, x_0) + \lim_{\widehat{g}^*(x_f,x_r,x_0)\to 1} Med(\varepsilon_{fi} \mid \widehat{\varepsilon}_i < \widehat{g}^*(x_f, x_r, x_0)) \\
&= g_f(x_f, x_0) + Med(\varepsilon_f \mid \widehat{\varepsilon} < 1) \\
&= g_f(x_f, x_0) + Med(\varepsilon_f) \\
&= g_f(x_f, x_0).
\end{aligned}$$

Thus we are done.

Another important point we want to make is that the model is not identified without identification at infinity.

To see why suppose that $\widehat{g}^*(X_{fi}, X_{ri}, X_{0i}))$ is bounded from above at g^u then if $\widehat{\varepsilon}_i > g^u$, $J_i = r$. Thus the data is completely uninformative about

$$E(Y_{fi} \mid \widehat{\varepsilon}_i > g^u)$$

so the model is not identified.

Parametric assumptions on the distribution of the error term is an alternative.

Really this is the same point as in the regression example we talk about to undergraduates-you can not predict outside the range of the data.

Step 3: Identification of g_r

What will be crucial is the other exclusion restriction (i.e. X_f).

Following an argument similar to in the previous case for any $\left(x_r^a, x_0^a\right)$ and $\left(x_r^b, x_0^b\right)$ suppose we want to identify $g_r\left(x_r^b, x_0^b\right) - g_r\left(x_r^a, x_0^a\right)$.

The key here is that the probability depends on X only through

$$g_f(x_f, x_0) - g_r(x_r, x_0)$$

and at this point g_f is known.

More formally we need to find x_f^a and x_f^b so that

$$Pr(J = f \mid X_f = x_f^a, X_r = x_r^a, X_0 = x_0^a) = Pr(J_i = f \mid X_f = x_f^b, X_r = x_r^b, X_0 = x_0^b)$$

But if this is the case it must be that

$$g_f(x_f^a, x_0^a) - g_r(x_r^a x_0^a) = g_f(x_f^b, x_0^b) - g_r(x_r^b x_0^b) \quad (1)$$

Rearranging equation (1) yields

$$g_r(x_r^b x_0^b) - g_r(x_r^a x_0^a) = g_f(x_f^b, x_0^b) - g_f(x_f^a, x_0^a)$$

Because $g_f(x_f^b, x_0^b) - g_f(x_f^a, x_0^a)$ has already been identified, $g_r(x_r^b x_0^b) - g_r(x_r^a x_0^a)$ is now identified also.

Step 4: Identification of G

There is an issue about location of g_r which is a hassle so lets ignore it.

To identify the joint distribution of $(\varepsilon_f, \widehat{\varepsilon}_r)$ note that from the data one can observe

$$\begin{aligned} & \Pr(J_i = f, \log(w_{fi}) < s \mid X_{0i} = x_0, X_{fi} = x_f, X_{ri} = x_r) \\ = \; & \Pr(g_f(x_f, x_0) + \varepsilon_{fi} > \widehat{g}_r(x_r, x_0) + \widehat{\varepsilon}_{ri}, g_f(x_f, x_0) + \varepsilon_{fi} < s) \\ = \; & \Pr(\varepsilon_{fi} - \widehat{\varepsilon}_{ri} < g_f(x_f, x_0) - \widehat{g}_r(x_r, x_0), \varepsilon_{fi} < s - g_f(x_f, x_0)) \end{aligned}$$

which is the cumulative distribution function of $(\varepsilon_{fi} - \widehat{\varepsilon}_{ri}, \varepsilon_{fi})$ evaluated at the point $(g_f(z_f, x) - \widehat{g}_r(z_r, x), s - g_f(z_f, x))$.

Thus the joint distribution of $(\varepsilon_{fi} - \widehat{\varepsilon}_{ri}, \varepsilon_{fi})$ is identified and getting from that to the joint distribution of $(\varepsilon_{ri}, \varepsilon_{fi})$ is straightforward.

Summary

The following table provides a summary.

Data Observed	Ass. on g_r, g_f	Ass. on G	Source
$J_i, Y_i, X_{0i}, X_{fi}, X_{ri}$	Supp(g_f, g_r) is $\mathbb{R}^2$	none	Heck/Honore
$J_i, Y_i \text{if} J_i = f$ X_{0i}, X_{fi}, X_{ri}	Supp (g_f, g_r) is $\mathbb{R}^2$	none	this chapter

An Example: Labor Supply

Here is what we plan to do

- Use PSID data
- Replicate Newey, Powell, Walker (1990) who estimate semi-parametric versions of some of Mroz's (1987) specifications
- Also show what happens if we try to identify the model using only normality for identification, but no exclusion restrictions

The Generalized Roy Model

Relax the assumption that choice only depends on income.

Let U_{fi} and U_{ri} be the utility that individual i would receive from being a fisherman or a hunter respectively where for $j \in \{f, r\}$,

$$U_{ji} = Y_{ji} + h_j(Z_i, X_{0i}) + \nu_{ji}.$$

The variable Z_i allows for the fact that there may be other variables that effect the taste for hunting versus fishing directly, but not affect wages in either sector. Workers choose to fish when

$$U_{fi} > U_{ri}.$$

We continue to assume that

$$\begin{aligned} Y_{fi} &= g_f(X_{fi}, X_{0i}) + \varepsilon_{fi} \\ Y_{ri} &= g_r(X_{ri}, X_{0i}) + \varepsilon_{ri}. \end{aligned}$$

The exact same proof can be used, just that step 2 is used separately for g_f and g_r

The only thing that is not identified is the joint distribution of $(\varepsilon_{fi}, \varepsilon_{ri})$ (Joint distribution of ν_i with each of these is identified)

Examples

There are a ton of examples of this model in labor economics:

1. Occupational choice
2. Schooling
3. Job Training
4. Migration
5. General treatment effects

Treatment Effects

There is a very large literature on the estimation of treatment effects.

We don't want to discuss full literature, but want to talk about how it fits into our framework

Assume the data is generated according to generalized Roy model

Define

$$\pi_i = Y_{fi} - Y_{ri}.$$

and think about identification of Average Treatment Effect

$$ATE \equiv E(\pi_i)$$

We focus on the “reduced form” selection model which here we define as $J_i = f$ when

$$h(Z_i) + \nu_i \geq 0$$

With the one additional assumption that expectations are finite, it is trivial to show that if the Generalized Roy model is identified, the ATE is identified.

It is weaker in that you only need an exclusion restriction in the selection equation, not in the outcome equation

It is really just “identification at infinity” (+ and -) with an exclusion restriction

For the most part the goal of different approaches in the treatment effect literature is to relax these assumptions in one way or another

Some focus on relaxing the support conditions

others focus on relaxing the exclusion restrictions.

First focus on relaxing the support conditions (and note that the authors don't necessarily sell it this way)

Local Average Treatment Effects

Imbens and Angrist (1994) essentially go to the other extreme and rather than assuming the exclusion restriction has full support, assume that it only takes on two values z^{ℓ} and z^{u}, in our framework they show that one can identify

$$E(\pi_i \mid h\left(z^{\ell}\right) \leq \nu_i < h(z^{h}))$$

which they call the local average treatment effect

Bounds on treatment Effects

Manski and others have focused on set identification rather than point identification.

Assume that $h(z^h)$ and $h(z^\ell)$ represent the upper and lower bounds of the support of $h(Z)$. Further assume that the support of Y_f and Y_r are bounded above by y^h and from below by y^ℓ. Then notice that

$$E(Y_f) = E(Y_f \mid J = f, Z = z^h) Pr(J = f \mid Z = z^h) + E(Y_f \mid \nu > h(z^h))(1 - Pr(J = f \mid Z = z^h))$$

We know everything here but $E(Y_f \mid \nu > h(z^h))$ which was exactly what we said couldn't be identified without identification at infinity

Putting this together with bounds on Y one can show that

$$
\begin{aligned}
&E(Y_f \mid J = f, Z = z^h)Pr(J = f \mid Z = z^h) + y^\ell(1 - Pr(J = f \mid Z = z^h)) \\
&\quad - E(Y_r \mid J = r, Z = z^\ell)Pr(J = r \mid Z = z^\ell) + y^u(1 - Pr(J = r \mid Z = z^\ell)) \\
&\leq ATE \leq \\
&E(Y_f \mid J = f, Z = z^h)Pr(J = f \mid Z = z^h) + y^u(1 - Pr(J = f \mid Z = z^h)) \\
&\quad - E(Y_r \mid J = r, Z = z^\ell)Pr(J = r \mid Z = z^\ell) + y^\ell(1 - Pr(J = r \mid Z = z^\ell)).
\end{aligned}
$$

Local Instrumental Variables and Marginal Treatment Effects

Heckman and Vytlacil (1999, 2001, 2005) construct a framework that is useful for constructing many types of treatment effects. They focus on the marginal treatment effect defined in our context as

$$\Delta^{MTE}(x_f, x_r, x_0, \nu) \equiv (E(\pi_i \mid (X_{fi}, X_{ri}, X_{0i}) = (x_f, x_r, x_0), \nu_i = \nu).$$

This is identified when Z_i is close to ν

They show that one can use this to build a lot of treatment effects if they are identified including the ATE

$$ATE = \int \int_0^1 \Delta^{MTE}(x_f, x_r, x_0, \nu) d\nu d\mu(x_f, x_r, x_0).$$

Relaxing the Exclusion Restriction Assumption

We next think of relaxing the exclusion restrictions.

We know of two main nonparametric alternatives in this case

Selection only on Observables

Assumption

ν is independent of $(\varepsilon_f, \varepsilon_r)$

Interestingly this is still not enough if there are values of observable covariates (X_f, X_r, X_0) for which $Pr(J = f \mid X_f, X_r, X_0) = 1$ or $Pr(J = f \mid X_f, X_r, X_0) = 0$

Thus we need the additional assumption

Assumption

For almost all (x_f, x_r, x_0) in the support of (X_f, X_r, X_0),

$$0 < Pr(J = f \mid X_f = x_f, X_r = x_r, X_0 = x_0) < 1$$

Theorem

Under assumptions 4.4 and 4.5 the Average Treatment Effect is identified

Estimation in this case is relatively straightforward. One can use matching or regression analysis to estimate the average treatment effect.

Set Estimates of Treatment Effects

There are quite a few ways to do this

- No assumption bounds
- Montone Treatment Effects
- Monotone Treatment Response
- Montone Selection
- Use selection on observables to bound selection on unobservables

Labor Supply Model with Continuous Hours Decision

The continuous selection model

We still have not decided exactly which model to use.

Duration Models and Search Models

Now consider models in which we observe $T_i > 0$ which is typically the length of time until some event occurs (like finding a job or death)

The most common set up is that T_i must be positive it is natural to model T_i using the basic framework we have been using all along:

$$log(T_i) = g(X_i) + \nu_i.$$

Clearly if we could observe the distribution of $log(T_i)$ conditional on X_i, identification of g and the distribution of ν_i would be straightforward, it is just a regression model

Competing Risk Model

However, typically we can not observed the full duration of T_i for some observations because something else happens to cut it short (data ends, guy dies, etc.)

Lets put this into our Roy model framework in that assume an unemployed worker will search until they find a job as a hunter or as a fisherman.

Write the model as:

$$log(T_{fi}) = g_f(X_{fi}, X_{0i}) + \nu_{fi}$$

$$log(T_{ri}) = g_f(X_{ri}, X_{0i}) + \nu_{ri}$$

Where the econometrician observes J_i and $\min\{T_{fi}, T_{ri}\}$

Note that this is just the Roy model with a minimum rather than a maximum.

All our results for the Roy model can be used here.

A common way to use duration data is to start with the mixed competing risk model

$$h(t \mid X_i) = \psi(t)\phi(X_i)\nu_i$$

where ψ is referred to as the baseline hazard, ν_i is an unobservable variable which is independent of the observables, and X_i is observable characteristics.

Heckman and Honore (1989) show how to map this into a framework similar to above.

In particular they show that one can write the models as

$$log(Z(T_i)) = g(X_i) + \nu_i^*.$$

where

- $g(\cdot) = -\log(phi(\cdot))$
- ν_i^* is a convolution of ν_i and an extreme value
- $Z(t)$ is the integrated baseline hazard

$$Z(t) \equiv \int_0^t \psi(t).$$

Heckman and Honore (1989) use a more general framework in which

$$S(t_f, t_r \mid X_i = x) = K(\exp\{-Z_f(t_f)\phi_f(x)\} \exp\{-Z_r(t_r)\phi_r(x)\}).$$

Theorem (Heckman and Honore, Theorem 1)

Assume that (T_f, T_r) *has the joint survivor function as given in (45). Then* $Z_f, Z_r, \phi_f, \phi_r,$ *and* K *are identified from the identified minimum of* (T_1, T_2) *under the following assumptions*

1. K *is continuously differentiable with partial derivatives* K_1 *and* K_2 *for* $i = 1, 2$ *the limit as* $n \to \infty$ *of* $K_i(\eta_{1n}, \eta_{2n})$ *is finite for all sequences of* η_{1n}, η_{2n} *for which* $\eta_{1n} \to 1$ *and* $\eta_{2n} \to 1$ *for* $n \to \infty$. *We also assume that* K *is strictly increasing in each of its arguments in all of* $[0, 1] \times [0, 1]$.
2. $Z_1(1) = 1, Z_2(1) = 1, \phi_1(x_0) = 1$ *and* $\phi_2(x_0) = 1$ *for some fixed point* x_0 *in the support* X.
3. *The support of* $\{\phi_1(x), \phi_2(x)\}$ *is* $(0, \infty) \times (0, \infty)$.
4. Z_1 *and* Z_2 *are nonnegative, differentiable, strictly increasing functions, except that we allow them to be* ∞ *for finite* t.

Note that they did not need an exclusion restriction

However, they still use something like identification at infinity: they look at the property of the model when T_{ji} is close to zero.

This would be less natural in a Roy type model

Search Models

Flinn and Heckman (1982) show the search model is not identified

The reservation wage rV is defined implicitly by the formula

$$c + rV = \frac{\lambda}{r}\int_{rV}^{\infty}(x - rV)dF(x)$$

where c is search cost, r is the interest rate, λ is the hazard rate of finding a job, and F is the offer arrival distribution

They assume that one observes the time until finding a job T_i and the wage a worker receives conditional on finding the job.

Identification:

- Reservation wage is identified trivially
- Distribution of wages above reservation trivially identified

$$\frac{f(x)}{1 - F(rV)} \text{ for } x \geq rV.$$

- hazard rates to job finding which is

$$\lambda(1 - F(rV)).$$

However, this is all that can be identified.

One can not separate λ from $(1 - F(rV))$ and one can not really say anything about f below rV

We combine this model with other assumptions to show that with exclusion restrictions it can be identified

The arrival rate of job offers is

$$\psi(t)\phi(X_{0i})\nu_i.$$

To be close to the Roy model above we assume the reservation wage can be written as

$$g_r(X_{ri}, X_{0i}) + \varepsilon_{ri}.$$

Finally we assume the wage offer that individual i would receive at time t is

$$\log(w_{fit}) = g_f(X_{fi}, X_{0i}) + \varepsilon_{fit}.$$

The complicated aspect of this model is that workers will potentially receive repeated offers that presumably will be correlated with each other.

We assume that the distribution of ε_{fit} is individual specific coming from distribution $F_{i\varepsilon_f}$.

That is each time a worker gets a new offer it is a draw from the distribution of $F_{i\varepsilon_f}$. As above (X_{fi}, X_{ri}, X_{0i}) is observable and independent of $(\nu_i, \varepsilon_{fi}, \varepsilon_{ri})$.

Theorem

Under Assumptions ???? Given that we observe T_i and w_{fiT_i} from the model determined by 50,50,51, we can identify ψ, ϕ, g_f, g_r, the joint distribution of $(\nu_i, \varepsilon_{fi}, \varepsilon_{ri})$ and the distribution of ξ_{it}.

Note that we can not get the full distribution of offer distributions.

This is not particularly surprising given the nature of the data

Panel data should be able to solve this problem.

Forward Looking Dynamic Models

We closely follow Taber (2000)

Now assume that one has to go through an internship before becoming a hunter.

After that decide whether to hunt deer or rabbits

When you make the internship decision you don't know which profession you will choose

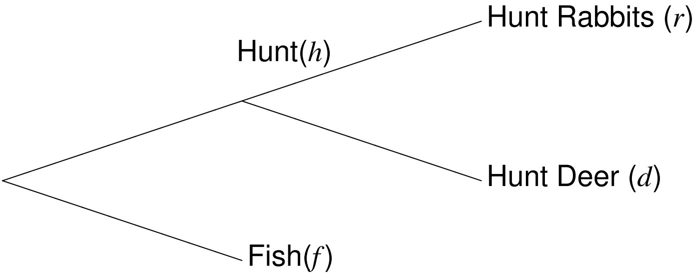
We would specify the model using the three value functions

$$V_{fi} = g_f(X_i) + \varepsilon_{fi}$$

$$V_{ri} = g_r(X_i) + \varepsilon_{ri}$$

$$V_{di} = 0$$

where V_{ji} is the relevant lifetime value function


Hunt Rabbits (r)
Hunt(h)
Hunt Deer (d)
Fish(f)

Known to the Agent at time one	Learned by the Agent at time two	Observed by Econo-metrician
$\varepsilon_{1i}, \varepsilon_{fi}$	ε_{ri}	X_{1i}, X_{fi}
X_{1i}, X_{fi}	X_{ri}	X_{ri}
$G(X_{ri} \mid X_{1i})$		J_i

Identification of g_a and g_b

Assumption

For any $x_r \in$supp$\{X_{ri}\}$ and $x_1 \in$supp$\{X_{1i}\}$,

$$\begin{aligned} \textit{supp}\{\varepsilon_{fi}\} &= (S^{\ell}_{\varepsilon_f}, S^{u}_{\varepsilon_f}) \subset \textit{supp}\{-g_f(X_{fi}) \mid X_{ri} = x_r)\} \\ \textit{supp}\{\varepsilon_{ri}\} &= (S^{\ell}_{\varepsilon_r}, S^{u}_{\varepsilon_r}) \end{aligned}$$

($S^{\ell}_{\varepsilon_f}, S^{u}_{\varepsilon_f}, S^{\ell}_{\varepsilon_r}$, and $S^{u}_{\varepsilon_r}$ need not be finite)

Assumption

For any $x_f \in$supp$\{X_{fi}\}$, $y \in (-S^{\ell}_{\varepsilon_r}, -S^{u}_{\varepsilon_r})$, and $c \in (0,1)$, there exists a set $\mathcal{X}_1(x_f, y, c)$ with positive measure such that for $x_1 \in \mathcal{X}_1(x_f, y, c)$,

(a) $x_f = X_f(x_1)$

(b) $\Pr(g_r < y \mid X_1 = x_1) > c$

(c) *The distribution of g_r conditional on x_1 is stochastically dominated by the unconditional distribution of g_r.*

Assumption

$(\varepsilon_{1i}, \varepsilon_{fi}, \varepsilon_{ri})$ *is independent of* (X_{1i}, X_{fi}, X_{ri}),

$$E(|\varepsilon_{ri}| \mid \varepsilon_{1i}) < \infty$$

and

$$E\left(|g_b\left(X_{ri}\right)| \mid X_{1i}\right) < \infty$$

Theorem

Under assumption 7.6,7.7, and 7.8, from data on $(X_{1i}, X_{fi}, X_{ri}, J_i)$ g_a *and* g_b *are identified up to monotonic transformations within* $(-S^u_{\varepsilon_f}, -S^\ell_{\varepsilon_f})$ *and* $(-S^u_{\varepsilon_r}, -S^\ell_{\varepsilon_r})$ *respectively.*

This is just a dynamic version of the Roy model discussed above

We need exclusion restriction that

- enters g_r but not g_f
- another that affects choices at time 1, does not enter g_f

Identification of Error Structure

Assumption

The outcome of the distribution of ε_{ri} conditional on ε_{1i} is the same as the agent's conditional expectation of this object.

Assumption

$\varepsilon_{ri} = \nu_{ri} + \eta_{ri}$ *where* $\nu_{ri} = E(\varepsilon_{ri} | \varepsilon_{1i})$, η_{ri} *is independent of* ε_{1i}, *and the characteristic functions of* ε_{ri} *and* η_{ri} *do not vanish.*

Assumption

$g_f(X_{fi})$ and $g_r(X_{ri})$ are identified.

Taber (2000) shows that these assumptions are not sufficient for identification without additional assumptions.

Assumption

For almost all $x_1 \in$supp(X_{1i}), $(S^{\ell}_{\varepsilon_r}, S^{u}_{\varepsilon_r}) \in$ supp$(-g_r(X_{ri}) \mid X_{1i} = x_1)$.

Theorem

Under Assumptions 7.6-7.10, the full model is identified

This last type of exclusion restriction is different

We need surprise, that is we need something that effects the expectation of g_b, but not g_a or g_b directly.

Equilibrium Models

Take the Roy model and add demand.

We think this is pretty straight forward.

Models with Learning

Talk about employer learning about ability

We are not sure the best papers to look at here

Signaling

Again a few scattered papers, perhaps we could tie these two together

Hedonic Models

Relate to IO literature here

First identification of Hedonic price equation

Recover Preferences from there

Measurement Error

Models such as

$$Y_{i1} = \theta_i + \nu_{i1}$$

$$Y_{i2} = \theta_i + \nu_{i2}$$

Identified from Kotlarski theorem.

Not directly related but pretty useful

Peer Effects

Reflection problem and stuff like that