The Generalized Roy Model and Treatment Effects

Christopher Taber

University of Wisconsin

November 10, 2016

Introduction

From Imbens and Angrist we showed that if one runs IV, we get estimates of the Local Average Treatment Effect

However, we are not limited to LATE, there are a lot of other things we could potentially do

In this set of lecture notes I want to put a little more structure on the model and think about what can and can not be identified under different assumptions

These models are the basis for a large class of structural models

We will start with the Roy model, then the Generalized Roy model, and then use that to think about identification of treatment effects

The Roy Model

Economy is a Village

There are two occupations

- hunter
- fisherman

Fish and Rabbits are going to be completely homogeneous

No uncertainty in number you catch

Hunting is “easier” you just set traps

Let

- π_F be the price of fish
- π_R be the price of rabbits
- F number of fish caught
- R number of rabbits caught

Wages are thus

$$\begin{aligned} W_F &= \pi_F F \\ W_R &= \pi_R R \end{aligned}$$

Each individual chooses the occupation with the highest wage

Thats it, that is the model

The relationship between this and the treatment effect framework should be transparent: this is a model in which people choose the treatment $t = 0$ or 1 to maximize Y_{ti}

This is an econometrics class rather than a labor class, so I really shouldn't talk about the model but I can't help myself

Key questions:

- Do the best hunters hunt?
- Do the best fisherman fish?

It turns out that the answer to this question depends on the variance of skill-nothing else

Whichever happens to have the largest variance in logs will tend to have more sorting.

In particular demand doesn't matter

To think of this graphically note that you are just indifferent between hunting and fishing when

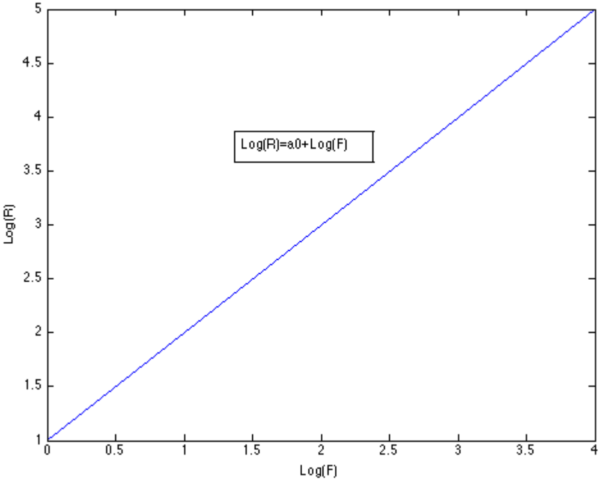
$$\log(\pi_R) + \log(R) = \log(\pi_F) + \log(F)$$

which can be written as

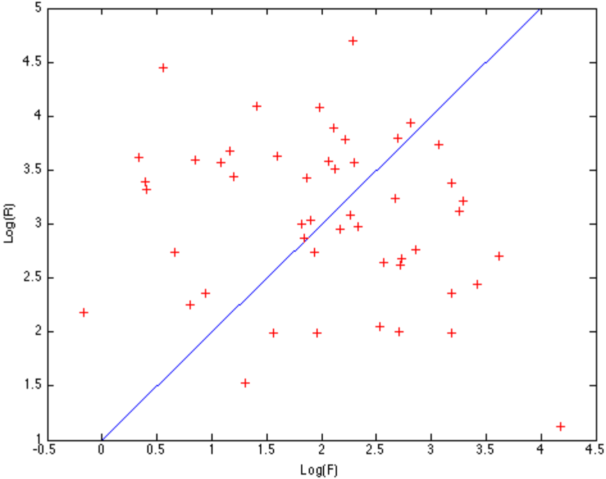
$$\log(R) = \log(\pi_F) - \log(\pi_R) + \log(F)$$

If you are above this line you hunt

If you are below it you fish



Log(R)=a0+Log(F)
Log(R)
Log(F)
5
4.5
4
3.5
3
2.5
2
1.5
1
0
0.5
1
1.5
2
2.5
3
3.5
4



5
4.5
4
3.5
3
2.5
2
1.5
1
Log(R)
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Log(F)

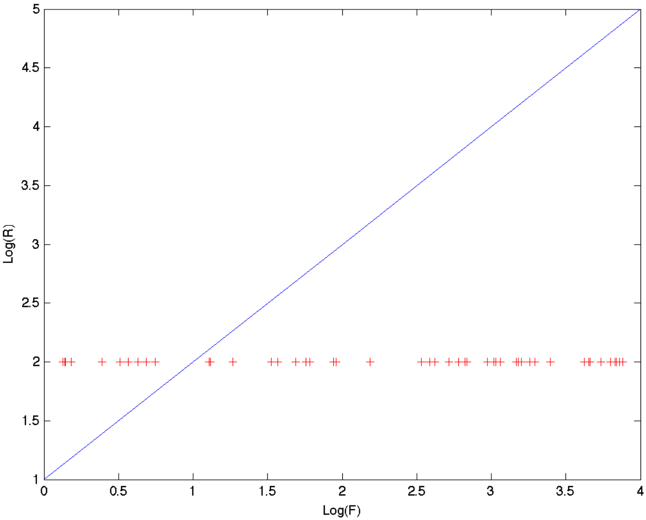
Case 1: No variance in Rabbits

Suppose everyone catches R^*

If you hunt you receive $W^* = \pi_R R^*$

Fish if $F > \frac{W^*}{\pi_F}$ Hunt if $F \leq \frac{W^*}{\pi_F}$

- The best fisherman fish
- All who fish make more than all who hunt



5
4.5
4
3.5
3
2.5
2
1.5
1
Log(R)
0
0.5
1
1.5
2
2.5
3
3.5
4
Log(F)

Case 2: Perfect correlation

Suppose that

$$\log(R) = \alpha_0 + \alpha_1 \log(F)$$

with $\alpha_1 > 0$

$$var(\log(R)) = \alpha_1^2 var(\log(F))$$

Thus if $\alpha_1 > 1$ then $var(\log(R)) > var(\log(F))$

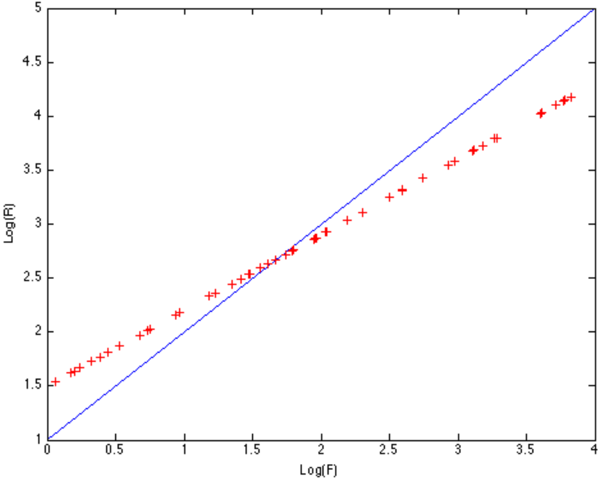
if $\alpha_1 < 1$ then $var(\log(R)) < var(\log(F))$

Fish if

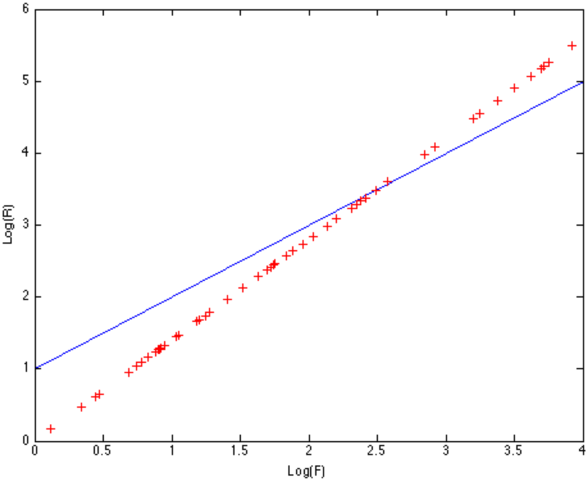
$$
\begin{aligned}
\log(W_F) &\geq \log(W_r) \\
\log(\pi_F) + \log(F) &\geq \log(\pi_R) + \log(R) \\
\log(\pi_F) + \log(F) &\geq \log(\pi_R) + \alpha_0 + \alpha_1 \log(F) \\
(1 - \alpha_1) \log(F) &\geq \log(\pi_R) + \alpha_0 - \log(\pi_F)
\end{aligned}
$$

If $\alpha_1 < 1$ then left hand side is increasing in $\log(F)$ meaning that better fisherman are more likely to fish

This also means that the best hunters fish


5
4.5
4
3.5
3
2.5
2
1.5
1
Log(R)
0
0.5
1
1.5
2
2.5
3
3.5
4
Log(F)

If $\alpha_1 > 1$ pattern reverses itself



Log(R)
Log(F)
0
1
2
3
4
5
6
0
0.5
1
1.5
2
2.5
3
3.5
4

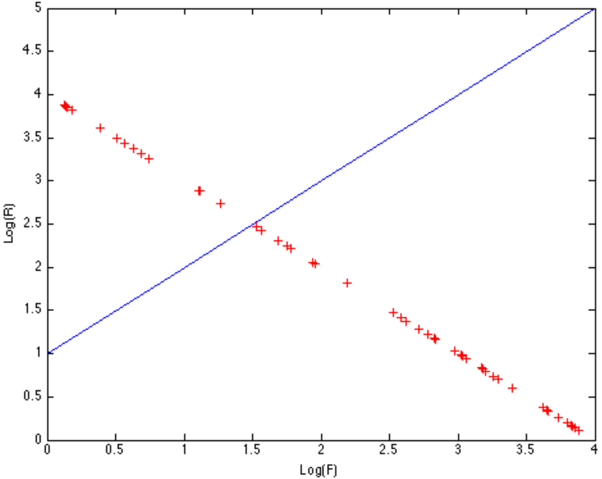
Thus when $var(\log(F)) > var(\log(R))$ (or $\alpha < 1$) the best fishermen fish, the best hunters fish

when $var(\log(R)) > var(\log(F))$ (or $\alpha > 1$) the best hunters hunt, the best fishermen hunt

Case 3: Perfect Negative Correlation

Exactly as before

$$(1 - \alpha_1) \log(F) \geq \log(\pi_R) + \alpha_0 - \log(\pi_F)$$


5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Log(R)
0
0.5
1
1.5
2
2.5
3
3.5
4
Log(F)

Best fisherman still fish

Best hunters hunt

Case 4: Log Normal Random Variables

This can all be formalized with Log normal random variables.

Survey papers can be found in various places but I won't get into it here.

Instead we focus on nonparametric identification of the Roy model.

Why is thinking about nonparametric identification useful?

- Speaking for myself, I think it is. I always begin a research project by thinking about nonparametric identification.
- Literature on nonparametric identification not particularly highly cited
- At the same time this literature has had a huge impact on empirical work in practice. A Heckman two step model without an exclusion restriction is often viewed as highly problematic these days- because of nonparametric identification
- It is also useful for telling you what questions the data can possibly answer. If what you are interested is not nonparametrically identified, it is not obvious you should proceed with what you are doing

Definition of Identification

Another term that means different things to different people

I will base my discussion on Matzkin's (2007) formal definition of identification but use my own notation and be a bit less formal

This will all be about the **Population** in thinking about identification we will completely ignore sampling issues

Data Generating Process

Let me define the data generating process in the following way

$$
\begin{aligned}
X_i \sim & H_0(X_i) \\
u_i \sim & F_0(u_i;\theta) \\
\Upsilon_i = & y_0(X_i, u_i;\theta)
\end{aligned}
$$

The data is (Υ_i, X_i) with u_i unobserved.

We know this model up to θ

To think of this as non-parametric we can think of θ as infinite dimensional

For example if F_0 is nonparametric we could write the model as $\theta = (\theta_1, F_0(\cdot))$

To put the Roy Model in this context we need to add some more structure to go from an economic model into an econometric model.

This means writing down the full data generation model.

First a normalization is in order.

We can redefine the units of F and R arbitrarily Lets normalize

$$\pi_F = \pi_R = 1$$

We consider the model

$$\begin{aligned} W_{fi} &= g_f(X_{fi}, X_{0i}) + \varepsilon_{fi} \\ W_{hi} &= g_h(X_{hi}, X_{0i}) + \varepsilon_{hi} \end{aligned}$$

where the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$ is G.

Let F_i be a dummy variable indicating whether the worker is a fisherman.

We can observe F_i and

$$W_i \equiv F_i W_{fi} + (1 - F_i) W_{hi}$$

Thus in this case

$$\begin{aligned}
\Upsilon_i &= (F_i, W_i) \\
X_i &= (X_{0i}, X_{fi}, X_{hi}) \\
u_i &= (\varepsilon_{fi}, \varepsilon_{hi}) \\
\theta &= (g_f, g_h, G) \\
y_0(X_i, u_i; \theta) &= \begin{bmatrix} 1\left(g_f(X_{fi}, X_{0i}) + \varepsilon_{fi} > g_h(X_{hi}, X_{0i}) + \varepsilon_{hi}\right) \\ \max\left\{g_f(X_{fi}, X_{0i}) + \varepsilon_{fi}, g_h(X_{hi}, X_{0i}) + \varepsilon_{hi}\right\} \end{bmatrix}
\end{aligned}$$

You can see the selection problem-we only observe the wage in the occupation the worker chose, we don't observe the wage in the occupation they didn't

Point Identification of the Model

The model is identified if there is a unique θ that could have generated the population distribution of the observable data (X_i, Υ_i)

A bit more formally, let Θ be the parameter space of θ and let θ_0 be the true value

- If there is some other $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$ for which the joint distribution of (X_i, Υ_i) when generated by θ_1 is identical to the joint distribution of (X_i, Υ_i) when generated by θ_0 then θ is not (point) identified
- If there is no such $\theta_1 \in \Theta$ then θ is (point) identified

Set Identification of the Model

Define Θ_I as the identified set.

I still want to think of there as being one true θ_0

Θ_I is the set of $\theta_1 \in \Theta$ for which the joint distribution of (X_i, Υ_i) when generated by θ_1 is identical to the joint distribution of (X_i, Υ_i) when generated by θ_0.

So another way to think about point identification is the case in which

$$\Theta_I = \{\theta_0\}$$

Identification of a feature of a model

Suppose we are interested not in the full model but only a feature of the model: $\psi(\theta)$

We can identify

$$\Psi_I \equiv \{\psi(\theta) : \theta \in \Theta_I\}$$

Most interesting cases occur when Θ_I is a large set but Ψ_I is a singleton

In practice $\psi(\theta)$ could be something complicated like a policy counterfactual in which we typically need to first get θ and then simulate $\psi(\theta)$

However, often it is much simpler and we can just write it as a known function of the data.

Classic example is the reduced form parameters in the simultaneous equations model-without an instrument the reduced form parameters are identified but the structural ones are not

Identification of the Roy Model

Lets think about identifying this model

The reference is Heckman and Honore (EMA, 1990)

I loosely follow the discussion in French and Taber, Handbook of Labor Economics, 2011

While the model is about the simplest in the world, identification is difficult

Assumptions

- $(\varepsilon_{fi}, \varepsilon_{hi})$ is independent of $X_i = (X_{0i}, X_{fi}, X_{hi})$ and continuously distributed
- Normalize $E(\varepsilon_{fi})$=0
  To see why this is a normalization we can always subtract $E(\varepsilon_{fi})$ from ε_{fi} and add it to $g_f(X_{fi}, X_{0i})$ making no difference in the model itself
- Normalize the median of $\varepsilon_{fi} - \varepsilon_{hi}$ to zero.
  A bit non-standard but we can always add the median of $\varepsilon_{fi} - \varepsilon_{hi}$ to ε_{hi} and subtract it from $g_h(X_{hi}, X_{0i})$
- $supp(g_f(X_{fi}, x_0), g_h(X_{hi}, x_0)) = \mathbb{R}^2$ for all $x_0 \in supp(X_{0i})$
- The econometrician observes F_i and observes the wage when $F_i = 1$

Step 1: Identification of Reduced Form 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

$$\begin{aligned}
Pr(F_i = 1 \mid X_i = x) &= Pr(g_h(x_h, x_0) + \varepsilon_{ih} \leq g_f(x_f, x_0) + \varepsilon_{if}) \\
&= Pr(\varepsilon_{ih} - \varepsilon_{if} \leq g_f(x_f, x_0) - g_h(x_h, x_0)) \\
&= G_{h-f}(g_f(x_f, x_0) - g_h(x_h, x_0)),
\end{aligned}$$

where G_{h-f} is the distribution function for $\varepsilon_{ih} - \varepsilon_{if}$

We can not separate $g_f(x_f, x_0) - g_h(x_h, x_0)$ from G_{h-f}, but we can identify the combination

This turns out to be useful

It means that we know that for any two values x_1 and x_2, if

$$Pr(F_i = 1 \mid X_i = (x_0^a, x_h^a, x_f^a)) = Pr(F_i = 1 \mid X_i = (x_0^b, x_h^b, x_f^b))$$

then

$$g_f(x_f^a, x_0^a) - g_h(x_h^a, x_0^a) = g_f(x_f^b, x_0^b) - g_h(x_h^b, x_0^b)$$

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 W_{fi} conditional on $(X_i = x, F_i = 1.)$

In particular we can identify

$$E(W_i \mid X_i = x, F_i = 1) = g_f(x_f, x_0) + E(\varepsilon_{if} \mid \varepsilon_{ih} - \varepsilon_{if} < g_f(x_f, x_0) - g_h(x_h, x_0)).$$

Lets think about identifying g_f up to location.

That is, for any $\left(x_f^a, x_0^a\right)$ and $\left(x_f^b, x_0^b\right)$ we want to identify

$$g_f(x_f^b, x_0^b) - g_f(x_f^a, x_0^a)$$

An exclusion restriction is key

Take x_h^b to be any number you want. From step 1 and from the support assumption we know that we can identify a x_h^a such that

$$Pr(F_i = 1 \mid X_i = (x_0^a, x_h^a, x_f^a)) = Pr(F_i = 1 \mid X_i = (x_0^b, x_h^b, x_f^b))$$

which means that

$$g_f(x_f^a, x_0^a) - g_h(x_h^a, x_0^a) = g_f(x_f^b, x_0^b) - g_h(x_h^b, x_0^b)$$

But then

$$
\begin{aligned}
& E(W_i \mid X_i = (x_0^a, x_h^a, x_f^a), F_i = 1) - E(W_i \mid X_i = (x_0^b, x_h^b, x_f^b), F_i = 1) \\
= & g_f(x_f^b, x_0^b) - g_f(x_f^a, x_0^a) \\
& + E(\varepsilon_{if} \mid \varepsilon_{ih} - \varepsilon_{if} < g_f(x_f^a, x_0^a) - g_h(x_h^a, x_0^a))) \\
& - E(\varepsilon_{if} \mid \varepsilon_{ih} - \varepsilon_{if} < g_f(x_f^b, x_0^b) - g_h(x_h^b, x_0^b)) \\
= & g_f(x_f^b, x_0^b) - g_f(x_f^a, x_0^a)
\end{aligned}
$$

Identification at Infinity

What about the location?

Notice that

$$
\begin{aligned}
&\lim_{g_h(x_h,x_0)\to-\infty;\left(x_0,x_f\right)\text{ fixed}} E(W_i \mid X_i = (x_0, x_h, x_f), F_i = 1) \\
&= g_f(x_f, x_0) \\
&+ \lim_{g_h(x_h,x_0)\to-\infty;\left(x_0,x_f\right)\text{ fixed}} E(\varepsilon_{fi} \mid \varepsilon_{ih} - \varepsilon_{if} < g_f(x_f, x_0) - g_h(x_h, x_0))) \\
&= g_f(x_f, x_0) + E(\varepsilon_{fi}) \\
&= g_f(x_f, x_0)
\end{aligned}
$$

Thus we are done

An important point is that the model is not identified without identification at infinity.

To see why suppose that $g_f(x_f, x_0) - g_h(x_h, x_0)$ is bounded from above at g^u then if $\varepsilon_{ih} - \varepsilon_{if} > g^u$, we know for sure that $F_i = 0$. Thus the data is completely uninformative about

$$E(\varepsilon_{fi} \mid \varepsilon_{ih} - \varepsilon_{if} > g^u)$$

so the model is not identified.

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

Who cares about Location?

Actually we do, a lot

- Without our intercept we know something about wage variation within fishing
- However we can not compare the wage level of fishing to the wage level of hunting
- This is what we need to construct treatment effects

Step 3: Identification of g_h

For any (x_h, x_0) we want to identify $g_h(x_h, x_0)$

What will be crucial is the other exclusion restriction (i.e. X_{fi}).

Again from step 1 and the other support condition, we know that can find an x_f so that

$$Pr(F_i = 1 \mid X_i = (x_0, x_h, x_f)) = 0.5.$$

This means that

$$0.5 = \Pr\left(\varepsilon_{hi} - \varepsilon_{fi} \leq g_f(x_f, x_0) - g_h(x_h, x_0)\right).$$

But the fact that $\varepsilon_{hi} - \varepsilon_{fi}$ has median zero implies that

$$g_h(x_h, x_0) = g_f(x_f, x_0).$$

Since g_f is identified, $g_f(x_f, x_0)$ is known, so $g_h(x_h, x_0)$ is identified from this expression.

Step 4: Identification of G

To identify the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$ note that from the data one can observe

$$\begin{aligned}
& \Pr(J_i = f, \log(W_i) < s \mid X_i = x) \\
= \; & \Pr(g_h(x_h, x_0) + \varepsilon_{hi} \leq g_f(x_f, x_0) + \varepsilon_{fi}, g_f(x_f, x_0) + \varepsilon_{fi} \leq s) \\
= \; & \Pr(\varepsilon_{hi} - \varepsilon_{fi} \leq g_f(x_f, x_0) - g_h(x_h, x_0), \varepsilon_{fi} \leq s - g_f(x_f, x_0))
\end{aligned}$$

which is the cumulative distribution function of $(\varepsilon_{hi} - \varepsilon_{fi}, \varepsilon_{fi})$ evaluated at the point $(g_f(x_f, x_0) - g_h(x_h, x_0), s - g_f(x_f, x_0))$

Thus by varying $(g_f(x_f, x_0) - g_h(x_h, x_0)$ and $(s - g_f(x_f, x_0))$ we can identify the joint distribution of $(\varepsilon_{hi} - \varepsilon_{fi}, \varepsilon_{fi})$

from this we can get the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$.

The Generalized Roy Model

Now we relax the assumption that choice only depends on income.

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

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

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

$$U_{fi} > U_{hi}.$$

Since all that matters is relative utility we can define

$$\varphi(Z_i, X_{0i}) \equiv \varphi_f(Z_i, X_{0i}) - \varphi_h(Z_i, X_{0i})$$
$$\nu_i \equiv \nu_{if} - \nu_{ih}$$

We continue to assume that

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

We can simplify further by writing the reduced form of the first stage using

$$\begin{aligned} \varphi^*(Z_i, X_{0i}, X_{fi}, X_{hi}) \equiv& \varphi(Z_i, X_{0i}) + g_f(X_{fi}, X_{0i}) - g_h(X_{hi}, X_{0i}) \\ \nu_i^* \equiv& -\nu_i - \varepsilon_{fi} + \varepsilon_{hi} \end{aligned}$$

so that

$$U_{fi} - U_{hi} = \varphi^*(Z_i, X_{0i}, X_{fi}, X_{hi}) - \nu_i^*$$

Lets think about identification of this model using the same 4 steps we used for the pure Roy model

A major difference, though, is now we assume that we also have data on Y_{hi}

Step 1

This is identical to what we did before, notice that

$$\begin{aligned} Pr(F_i = 1 \mid X_i = x) &= Pr(U_{hi} \leq U_{fi}) \\ &= Pr(\nu_i^* \leq \varphi^*(Z_i, X_{0i}, X_{fi}, X_{hi})) \\ &= G_{\nu^*}(\varphi^*(Z_i, X_{0i}, X_{fi}, X_{hi})), \end{aligned}$$

as before we can not separate G_{ν^*} from $\varphi^*(Z_i, X_{0i}, X_{fi}, X_{hi})$, but we can identify level sets of $\varphi^*(Z_i, X_{0i}, X_{fi}, X_{hi})$

Step 2

Getting g_f is exactly the same as before

$$\lim_{\varphi^*(.)\to\infty;\left(x_0,x_f\right)\text{ fixed}} E(W_i \mid (Z_i, X_i) = (z, x_0, x_h, x_f), F_i = 1))$$

$$= g_f(x_f, x_0)$$

$$+ \lim_{\varphi^*(.)\to\infty;\left(x_0,x_f\right)\text{ fixed}} E(\varepsilon_{fi} \mid \nu_i^* < \varphi^*(z, x_0, x_f, x_h))$$

$$= g_f(x_f, x_0) + E(\varepsilon_{fi})$$

$$= g_f(x_f, x_0)$$

Step $2'$

Now think about identifying g_h-this is completely analogous to g_f

$$
\begin{aligned}
&\lim_{\varphi^*(.)\to-\infty;(x_0,x_h)\text{ fixed}} E(W_i \mid (Z_i, X_i) = (z, x_0, x_h, x_f), F_i = 0) \\
&= g_h(x_h, x_0) \\
&+ \lim_{\varphi^*(.)\to\infty;(x_0,x_h)\text{ fixed}} E(\varepsilon_{hi} \mid \nu_i^* > \varphi^*(z, x_0, x_f, x_h)) \\
&= g_h(x_h, x_0) + E(\varepsilon_{hi}) \\
&= g_h(x_h, x_0)
\end{aligned}
$$

Step 3

Again almost completely analogous to the case before since we can write

$$U_{fi} - U_{hi} = g_f(X_{fi}, X_{0i}) + \varphi(Z_i, X_{0i}) - g_h(X_{hi}, X_{0i}) \\ + \varepsilon_{fi} + \nu_i - \varepsilon_{hi}$$

We need exclusion restrictions in X_{fi} **or** X_{hi} since under the assumption that the median of $\varepsilon_{fi} + \nu_{fi} - \varepsilon_{hi} - \nu_{hi}$, means that if

$$Pr(J_i = f \mid X_i = x, Z_i = z) = 0.5$$

then

$$\varphi(z, x_0) = g_h(x_h, x_0) - g_f(x_f, x_0)$$

Step 4

Finally we need to identify the distribution of the error terms.

Recall that the model is

$$
\begin{aligned}
U_{fi} - U_{hi} &= \varphi^*(Z_i, X_i) - \nu_i^* \\
Y_{fi} &= g_f(X_{fi}, X_{0i}) + \varepsilon_{fi} \\
Y_{hi} &= g_h(X_{hi}, X_{0i}) + \varepsilon_{hi}
\end{aligned}
$$

We can use an argument identical to the standard Roy model

$$
\begin{aligned}
& Pr(F_i = 1, Y_{fi} \leq y \mid (Z_i, X_i) = (z, x)) \\
&\quad = Pr(\nu_i^* \leq \varphi^*(z, x), g_f(x_f, x_0) + \varepsilon_{fi} \leq y) \\
&\quad = G_{\nu^*, \varepsilon_f}\left(\varphi^*(z, x), y - g_f(x_f, x_0)\right).
\end{aligned}
$$

where G_{v,ε_f} is the joint distribution of $(\varepsilon_{fi} + \nu_i - \varepsilon_{hi}, \varepsilon_{fi})$.

Using an analogous argument we can get the joint distribution of $(\varepsilon_{fi} + \nu_i - \varepsilon_{hi}, \varepsilon_{hi})$

However, this is all that can be identified.

In particular the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$ is not identified.

The problem is that we never observe the same person as both a hunter and a fisherman

Even with Normal error terms the covariance between ε_{fi} and ε_{hi} is not identified

In the Roy model it was identified because from the choice/wage equation we identified the joint distribution of $(\varepsilon_{fi} - \varepsilon_{hi}, \varepsilon_{hi})$ from which you can get the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$

Clearly one can not do this from the joint distribution of $(\varepsilon_{fi} + \nu_i - \varepsilon_{hi}, \varepsilon_{hi})$.

Treatment Effects

The generalized Roy model can also be thought of as the "treatment effect model"

We just interpret Y_{fi} to be the outcome with treatment, Y_{hi} to be the outcome without treatment

This literature does not typically worry about the effect of the treatment on enrollment, so we can used the"reduced form" version of the selection model (I am also going to ignore other regressors for now, so the only observable is the exclusion restriction)

Thus the model is $J_i = f$ when

$$\varphi^*(Z_i) - \nu_i^* \geq 0$$

and we observe Y_{fi} when $F_i = 1$ and Y_{hi} when $F_i = 0$

Identification of Average Treatment Effect

Analogous to the previous set of lecture notes we define

$$\alpha_i = Y_{fi} - Y_{hi}.$$

Think about identification of Average Treatment Effect

$$\begin{aligned} ATE &\equiv E(\alpha_i) \\ &= E(Y_{fi}) - E(Y_{hi}). \end{aligned}$$

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

$$\begin{aligned}
&\lim_{\varphi^*(z)\to\infty} E(Y_i \mid F_i = 1, Z_i = z) - \lim_{\varphi^*(z)\to-\infty} E(Y_i \mid F_i = 0, Z_i = z) \\
&\qquad = E(Y_{fi}) - E(Y_{hi}) \\
&\qquad = ATE
\end{aligned}$$

There are two major limitations of this.

- We can identify the marginal distributions of Y_{fi} and Y_{hi} but not their joint distribution
- Without full support we can do even less

Lets think about each of these in turn

We showed that we can identify the joint distribution of $(\nu_i^*, \varepsilon_{fi})$. This means we can identify the marginal distribution of (ε_{fi}) and thus the marginal distribution of Y_{fi}

Analogously from the joint distribution of $(\varepsilon_{fi} + \nu_i - \varepsilon_{hi}, \varepsilon_{hi})$ we can identify the marginal distribution of of (ε_{hi}) and thus the marginal distribution of Y_{hi}.

However, we can not identify their joint distribution which limits what we can do.

It is important to understand quantile treatment effects within this context.

Quantile Treatment Effects

Lets assume the the support conditions and assume from identification at infinite we can identify the full marginal distribution of Y_{fi} and Y_{hi}.

When we estimate the Average Treatment Effect we compare

$$\widehat{ATE} = \overline{Y}_f - \overline{Y}_h$$

But there is nothing special about means, we could do the same thing with medians

$$\widehat{TE}_{[0.5]} = Y_{f[0.5]} - Y_{h[0.5]}$$

That is we can take the difference between the medians of the controls and the median of the treatments.

There is nothing special about the median. We can do this at any quantile

$$Pr(Y_i \leq Y_{[q]}) = q$$

$$\widehat{TE}_{[q]} = Y_{f[q]} - Y_{h[q]}$$

This is referred to as the quantile treatment effect

One important clarification

The quantile treatment effect is **NOT** the quantile of the treatment effects

For expectations

$$E\left(Y_{1i} - Y_{0i}\right) = E\left(Y_{1i}\right) - E\left(Y_{0i}\right)$$

Because the difference in the expectations is the expectation of the difference

This is not true with Quantiles-the difference in the medians is not the median of the difference

Example

Suppose only three people

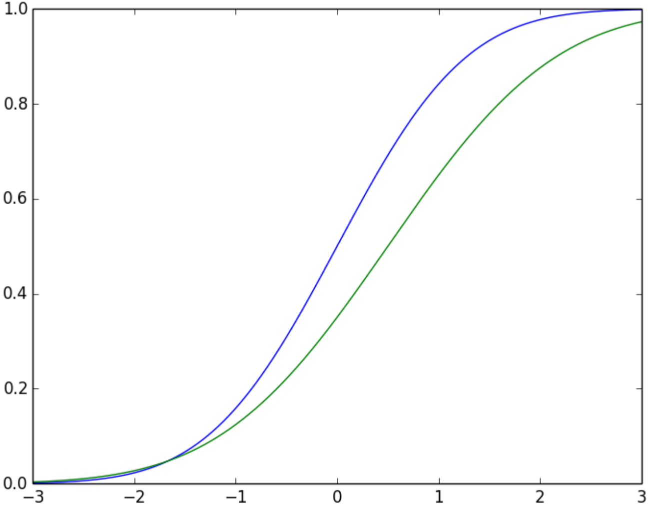
Person	Y_{hi}	Y_{fi}	Δ_i
1	1.2	3.4	2.2
2	2.3	2.8	0.5
3	4.4	5.1	0.7
Median	2.3	3.4	0.7

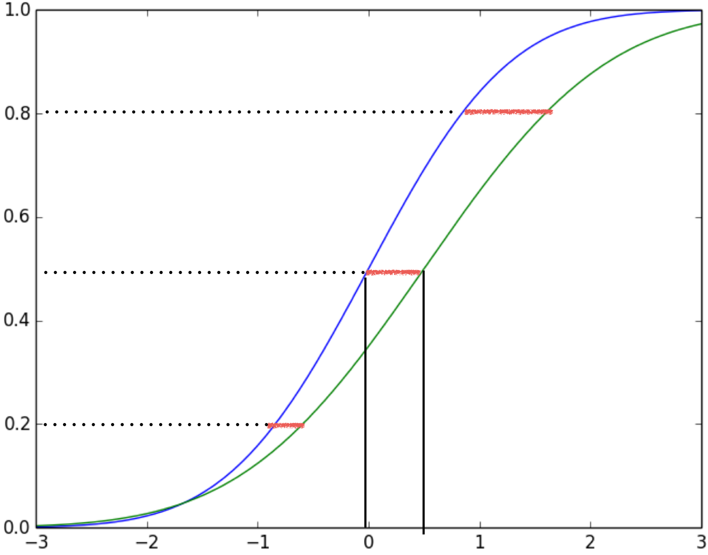
The median treatment effect is 3.4-2.3=1.1

The median of the treatment effect is 0.7

We cannot figure out what the median of the treatment effect is in general because we don't know who goes with who

The data identifies the two marginal distributions of Y_{fi} and Y_{hi} but in the generalized Roy model, it does not identify the joint distribution.


1.0
0.8
0.6
0.4
0.2
0.0
−3
−2
−1
0
1
2
3



1.0
0.8
0.6
0.4
0.2
0.0
−3
−2
−1
0
1
2
3

Social Welfare Function

The fact that we can not identify the joint distribution is not necessarily a problem

Suppose we want to compare the world with the treatment versus the world without it (not so interesting in pure Roy model).

If we have social welfare function $V(\cdot)$ we can calculate $E[V(Y_{fi})]$ and $E[V(Y_{hi})]$ from the marginal distributions only

Support Conditions

The second issue was the support conditions.

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-though this requires strong assumptions like no selection on unobservables

Lets think about literatures that think about relaxing the support conditions (and note that the authors don't usually sell it this way)

LATE

Imbens and Angrist showed that IV with a binary instrument converged to

$$E(\alpha_i \mid G_i = 3)$$

To map this into the current framework recall that we thought about the case in which Z_i is either one or zero

People choose the treatment (fish) when

$$\varphi^*(Z_i) - \nu_i^* \geq 0$$

The G_i people are those who would choose to fish when $Z_i = 1$ but not when $Z_i = 0$ This means that for them:

$$\varphi^*(1) - \nu_i^* \geq 0$$
$$\varphi^*(0) - \nu_i^* < 0$$

and thus

$$\varphi^*(0) < \nu_i^* \leq \varphi^*(1)$$

In other words the local average treatment effect can be written as

$$E(\alpha_i \mid \varphi^*(0) < \nu_i^* \leq \varphi^*(1))$$

Lets think about how this relates to identification at infinite and support considerations

Instead of binary suppose that Z_i had larger support.

For any two points in the support $z_1$and z_0 with $\varphi^*(z_1) > \varphi^*(z_0)$ we can use IV to identify

$$E(\alpha_i \mid \varphi^*(z_0) < \nu_i^* \leq \varphi^*(z_1))$$

Thus if I can find points so that $\varphi^*(z_0) = -\infty$ and $\varphi^*(z_1) = \infty$, then one can identify

$$E(\pi_i \mid -\infty < \nu_i^* \leq \infty) = ATE.$$

However if the support of $\varphi(Z_i)$ is bounded away from 0 or from 1, then this does not work

For example let $\mathcal{Z}$ be the support of Z_i.

Define

$$z^{\ell} \equiv \arg\min_{z \in \mathcal{Z}} \varphi^*(z)$$
$$z^{u} \equiv \arg\max_{z \in \mathcal{Z}} \varphi^*(z)$$

and if $z^{\ell} < 0$ then there is no state of the world in which an individual with $\nu_i < \varphi^*(z^{\ell})$ would ever hunt so we can never hope to identify the effect for them.

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 and showing what can be identified.

They focus on the marginal treatment effect defined in our context as

$$\Delta^{MTE}(x,\nu) \equiv E(\alpha_i \mid X_i = x, \nu_i = \nu).$$

To see where identification comes from, note that adding regressors to the argument above, it is easy to show that one can identify

$$E(\alpha_i \mid \varphi^*(z_0,x) < \nu_i^* \leq \varphi^*(z_1,x), X_i = x)$$

for any z_0 and z_1.

From this note that

$$
\begin{aligned}
\lim_{\varphi^*(z_0,x)\uparrow\nu,\varphi^*(z_1,x)\downarrow\nu} & E(\alpha_i \mid \varphi^*(z_0,x) < \nu_i^* \leq \varphi^*(z_1,x), X_i = x) \\
&= E(\alpha_i \mid X_i = x, \nu_i^* = \nu) \\
&= \Delta^{MTE}(x,\nu)
\end{aligned}
$$

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

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

They also show the instrumental variables estimator can be written as

$$\int_0^1 \Delta^{MTE}(x, \nu) h_{IV}(x, \nu) d\nu$$

for some function h_{IV} that they can calculate

Further they show how to use this framework to think about policy evaluation.

Bounds on treatment Effects

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

That is even if I can not point identify ATE, I can Identify a set in which ATE lies

As an example consider the instrument case in which $\varphi^*(z^h)$ and $\varphi^*(z^\ell)$ represent the upper and lower bounds of the support of $\varphi^*(Z_i)$.

Then notice that

$$
\begin{aligned}
E\left(Y_{fi}\right) = & E(Y_{fi} \mid F_i = 1, Z_i = z^h) Pr(F_i = 1 \mid Z = z^h) \\
& + E(Y_{fi} \mid \nu_i^* > \varphi^*(z^h))(1 - Pr(F_i = 1 \mid Z = z^h))
\end{aligned}
$$

We know everything here but $E(Y_{fi} \mid \nu_i > \varphi^*(z^h))$ which was exactly what we said couldn't be identified without identification at infinity

However suppose that the support of Y_{fi} and Y_{hi} are bounded above by y^h and from below by y^ℓ.

We know that

$$y^\ell \leq E(Y_{fi} \mid \nu_i^* > \varphi^*(z^h)) \leq y^h$$

Using this, one can show that

$$
\begin{aligned}
& E(Y_{fi} \mid F_i = 1, Z_i = z^h)P(z^h) + y^\ell(1 - P(z^h)) \\
& \qquad - E(Y_{hi} \mid F_i = 1, Z_i = z^\ell)(1 - P(z^\ell)) + y^u P(z^\ell) \\
& \quad \leq ATE \leq \\
& E(Y_{fi} \mid F_i = 1, Z_i = z^h)P(z^h) + y^u(1 - P(z^h)) \\
& \qquad - E(Y_{hi} \mid F_i = 0, Z_i = z^\ell)(1 - P(z^\ell)) + y^\ell P(z^\ell).
\end{aligned}
$$

where

$$P(z) \equiv Pr(F_i = 1 \mid Z_i = z)$$

Set Estimates of Treatment Effects

There are quite a few other ways to bound treatment effects

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