The Roy Model

Christopher Taber

Wisconsin

August 29, 2015

Outline

Basic Model

Implications of Model

Normal Random Variables

Identification

Parametric Roy Model

Nonparametric Identification of Roy Model

Outline

Basic Model

Implications of Model

Normal Random Variables

Identification

Parametric Roy Model

Nonparametric Identification of Roy Model

Economy is a Village

The Roy Model

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

Outline

Basic Model

Implications of Model

Normal Random Variables

Identification

Parametric Roy Model

Nonparametric Identification of Roy Model

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 grahically note that you are just indifferent between hunting and fishing when

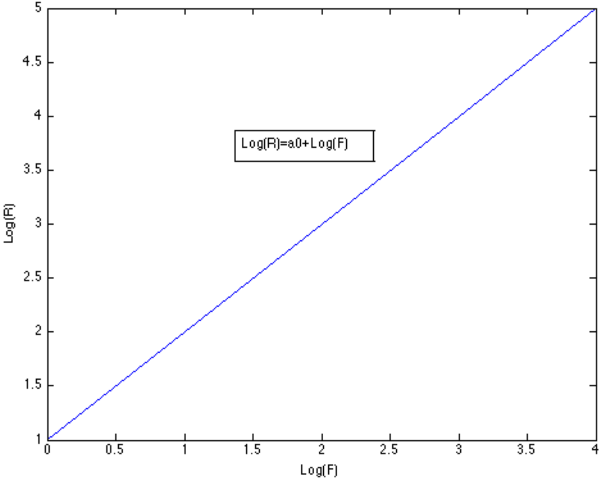
$$\log(\pi_R) + \log(R) = \log(\pi_F) + \log(F)$$

which can be written as

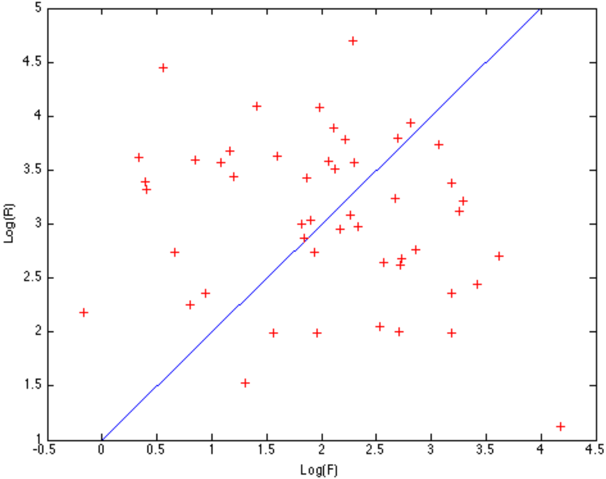
$$\log(R) = \log(\pi_F) - \log(\pi_R) + \log(F)$$

If you are above this line you hunt

If you are below it you fish



$$a_0 = \log(\pi_F) - \log(\pi_R)$$


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

Case 1: No variance in Rabbits

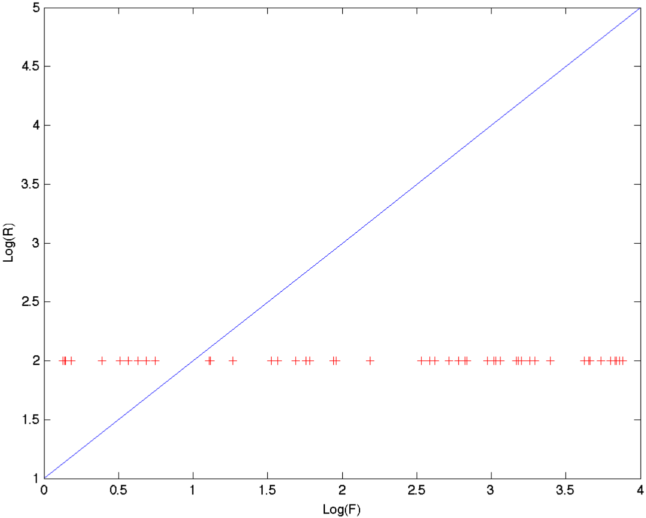
Suppose everyone catches R^*

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 $\alpha_1 > 0$

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

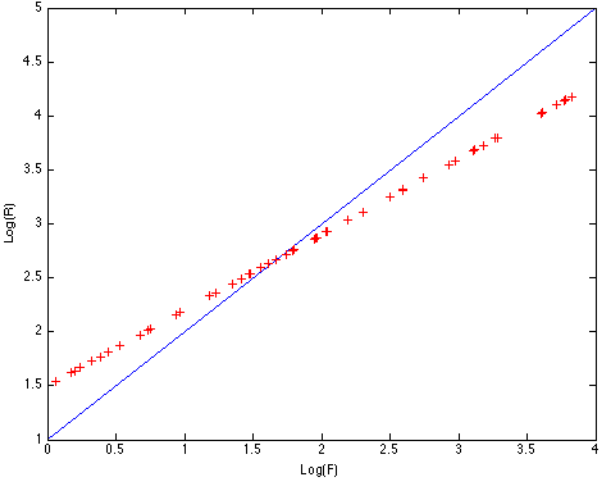
Fish if

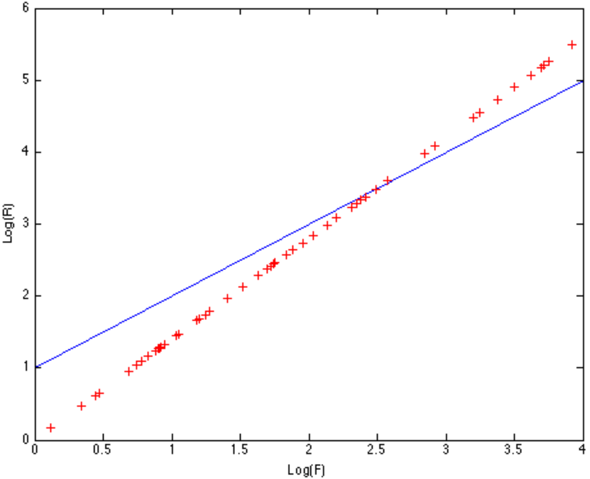
$$\begin{aligned}
\log(W_F) &\geq \log(W_r) \\
\log(\pi_F) + \log(F) &\geq \log(\pi_R) + \log(R) \\
\log(\pi_F) + \log(F) &\geq \log(\pi_R) + \alpha_0 + \alpha_1 \log(F) \\
(1 - \alpha_1)\log(F) &\geq \log(\pi_R) + \alpha_0 - \log(\pi_F)
\end{aligned}$$

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

This also means that the best hunters fish

If $\alpha_1 > 1$ pattern reverses itself

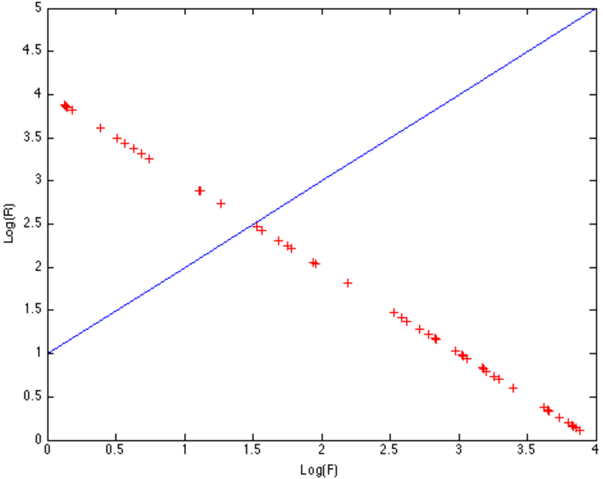

5
4.5
4
3.5
3
2.5
2
1.5
1
Log(R)
0
0.5
1
1.5
2
2.5
3
3.5
4
Log(F)


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

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

Lets try to formalize all of this

assume that

$$(\log(R), \log(F)) \sim N(\mu, \Sigma)$$

where

$$\begin{aligned} \mu &= \begin{bmatrix} \mu_F \\ \mu_R \end{bmatrix} \\ \Sigma &= \begin{bmatrix} \sigma_{FF} & \sigma_{RF} \\ \sigma_{RF} & \sigma_{RR} \end{bmatrix} \end{aligned}$$

Outline

Basic Model

Implications of Model

Normal Random Variables

Identification

Parametric Roy Model

Nonparametric Identification of Roy Model

Normal Random Variables

Lets stop for a second and review some properties of normal random variables

- Sum of Normals is Normal
- Described by first and second moments perfectly
- If $u \sim N(0,1)$then

$$\begin{aligned} E(u \mid u > k) &= \frac{\phi(k)}{1 - \Phi(k)} \\ &\equiv \lambda(-k) \end{aligned}$$

the inverse mills ratio

Putting the first two together there is a regression interpretation

Take any two normal variables (u_1, u_2) we can write

$$u_2 = \alpha_0 + \alpha_1 u_1 + \xi$$

as a regression with ξ normally distributed with 0 mean and independent of u_1

Notice that by definition

$$\begin{aligned} cov(u_1, u_2) &= cov(u_1, \alpha_0 + \alpha_1 u_1 + \xi) \\ &= \alpha_1 var(u_1) \end{aligned}$$

Therefore

$$\begin{aligned} \alpha_1 &= \frac{cov(u_1, u_2)}{var(u_1)} \\ \alpha_0 &= E(u_2) - \alpha_1 E(u_1) \end{aligned}$$

This last thing is the basis of the Heckman Two step

Suppose that

$$\begin{aligned} Y_1^* &= X'\beta + u_1 \\ Y_2 &= Z'\gamma + u_2 \end{aligned}$$

we observe d which is one if $Y_1^* > 0$ and zero otherwise.

Assume first part is a probit so $u_1 \sim N(0,1)$

So

$$Y_2 = Z'\gamma + \alpha_0 + \alpha_1 u_1 + \xi$$

Furthermore $\alpha_0 = 0$ if $E(u_1) = E(u_2) = 0$

Then

$$\begin{aligned}
&E(Y_2 \mid X, Z, d = 1) \\
&\quad = E(Y_2 \mid X, Z, u_1 = -X'\beta) \\
&\quad = Z'\gamma + \alpha_1 E(u_1 \mid X, Z, u_1 = -X'\beta) + E(\xi \mid X, Z, u_1 = -X'\beta) \\
&\quad = Z'\gamma + \alpha_1 \lambda(X'\beta)
\end{aligned}$$

Back to the Roy Model

Lets use this idea for the Roy model

Fish if

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

The question is what is

$$E\left(\log(\pi_F) + \log(F) \mid \log(\pi_F) + \log(F) > \log(\pi_R) + \log(R)\right)?$$

If it is bigger than $\log(\pi_F) + \mu_F$ then best fishermen fish (on average)

For $j \in \{R, F\}$,let

$$\begin{aligned} a_j &= \log(\pi_j) + \mu_j \\ u_j &= \log(j) - \mu_j \end{aligned}$$

Then

$$\begin{aligned} & E(\log(\pi_F) + \log(F) \mid \log(\pi_F) + \log(F) > \log(\pi_R) + \log(R)) \\ = \;& E(a_F + u_F \mid a_F + u_f > a_R + u_R) \\ = \;& a_f + E(u_F \mid u_F - u_R > a_R - a_F) \end{aligned}$$

Now think of the regression of u_F on $u_F - u_R$

$$u_F = \alpha \left(u_F - u_R\right) + \omega$$

where

$$\begin{aligned} \alpha &= \frac{cov(u_F, u_F - u_R)}{var(u_F - u_R)} \\ &= \frac{\sigma_{FF} - \sigma_{FR}}{\sigma^2} \end{aligned}$$

$$\sigma^2 = var(u_F - u_R)$$

So

$$\begin{aligned}
E\left(u_F \mid u_F - u_R > a_R - a_F\right) =& E\left(\alpha\left(u_F - u_R\right) + \omega \mid u_F - u_R > a_R - a_F\right. \\
=& \alpha\sigma E\left(\frac{u_F - u_R}{\sigma} \mid \frac{u_F - u_R}{\sigma} > \frac{a_R - a_F}{\sigma}\right) \\
=& \alpha\sigma\lambda\left(\frac{a_F - a_R}{\sigma}\right) \\
=& \frac{\sigma_{FF} - \sigma_{FR}}{\sigma}\lambda\left(\frac{a_F - a_R}{\sigma}\right)
\end{aligned}$$

The question boils down to the sign of this object.

If it is positive then positive selection into fishing

But $\sigma > 0$ and $\lambda\left(\right) > 0,$so the question is about the sign of

$$\sigma_{FF} - \sigma_{FR}$$

Notice that

$$
\begin{aligned}
Var(\log(F) - \log(R)) &= \sigma_{FF} + \sigma_{RR} - 2\sigma_{FR} \\
&= [\sigma_{FF} - \sigma_{FR}] + [\sigma_{RR} - \sigma_{FR}] \\
&> 0
\end{aligned}
$$

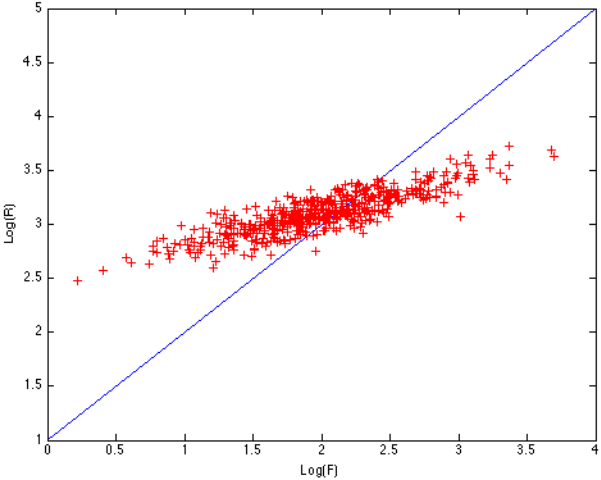
One of these must be positive.

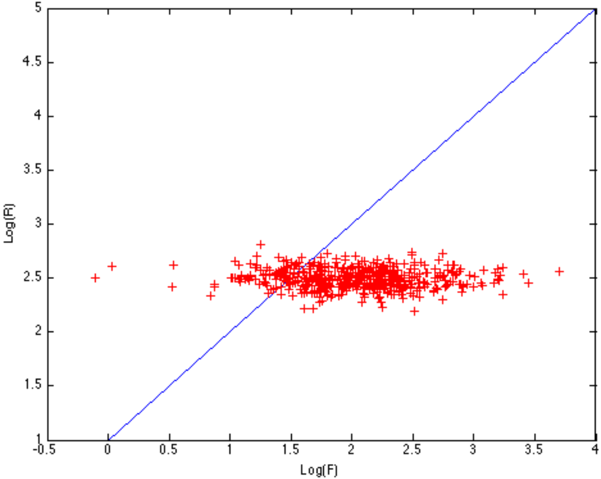
Thus if $\sigma_{FF} > \sigma_{RR}$ there is positive selection into fishing

Hunters could go either way depending on $\sigma_{RR} - \sigma_{FR}$

- If covariance is negative or zero, positive selection into hunting
- If correlation between the two is high enough, selection is negative
  In particular if they are perfectly correlated

$$\sigma_{FR} = \sqrt{\sigma_{RR}}\sqrt{\sigma_{FF}} > \sigma_{RR}$$


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)


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
Log(F)

Outline

Basic Model

Implications of Model

Normal Random Variables

Identification

Parametric Roy Model

Nonparametric Identification of Roy Model

How do we think about identifying this model?

This is discussed in Heckman and Honore (EMA, 1990)

We will follow the discussion of this in French and Taber (HLW, 2011) fairly closely

Lets think about how we would estimate the model

Suppose we have data on Occupation and Wages from a cross section

with

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

Can we identify $G(F, R)$- the joint distribution of F and R?

First a normalization is in order.

We can redefine the units of F and R arbitrarily

Lets normalize

$$\pi_F = \pi_R = 1$$

This still isn't enough in general

From the data we can observe

$$G(R \mid R > F)$$
$$G(F \mid R \leq F)$$

Lets think more generally about what identification means

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-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
- It is 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

We follow Matzkin's (2007) formal definition of identification and follow her notation exactly

Let ς represent a model or data generating process. It is essentially a combination of parameters, functions, and distribution functions where S is the space of functions in which ς lies.

As an example consider the semiparametric regression model

$$Y_i = X_i'\beta + \varepsilon_i$$

with

$$E(\varepsilon_i \mid X_i) = 0$$

In this case $\varsigma = (\beta, F_{X,\varepsilon})$ where $F_{X,\varepsilon}$ is the joint distribution between X_i and ε_i

S is the set of permissible β and $F_{X,\varepsilon}$

The data we can potentially observe is the full joint distribution of (Y_i, X_i)

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* $\psi \in \Omega$,

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

So what the heck does that mean?

Basically $\Psi(\varsigma)$. Measures some feature of the model.

Interesting examples in our case are:

- $\Psi(\varsigma) = \varsigma$
- $\Psi(\varsigma) = \beta$
- $\Psi(\varsigma) =$ the effect of some policy counterfactual

From this, $\Gamma_{Y,X}(\psi, S)$ is the set of possible data distributions that are consistent with the model and a given value $\Psi(\varsigma) = \psi$

$\psi^* \in \Omega$ is identified when there is no other value of ψ that is consistent with the joint distribution of the data

Outline

Basic Model

Implications of Model

Normal Random Variables

Identification

Parametric Roy Model

Nonparametric Identification of Roy Model

Before thinking about nonparametric identification, lets think about parametric estimation

If you understand that, it will turn out that the nonparametric identification is analogous.

French and Taber focus on the labor supply case, and we will as well

That is let

$$
\begin{gathered}
Y_{fi} = X_{0i}'\gamma_{0f} + X_{fi}'\gamma_{ff} + \varepsilon_{fi} \\
Y_{hi} = X_{0i}'\gamma_{0h} + X_{hi}'\gamma_{hh} + \varepsilon_{hi} \\
\begin{bmatrix} \varepsilon_{fi} \\ \varepsilon_{hi} \end{bmatrix} = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_f^2 & \sigma_{fh} \\ \sigma_{fh} & \sigma_h^2 \end{bmatrix} \right).
\end{gathered}
$$

But we will take Y_{fi} to be market production and Y_{hi} to be market production so the individual works if $Y_{fi} > Y_{hi}$

The econometrician gets to observe whether the individual works, and if they work they observe the wage

The key distinction between this and the more general Roy model is that the econometrician does not observe Y_{hi} for people who do not work (which seems reasonable in the labor supply problem)

We can estimate this model in 4 steps:

We could also just estimate the whole thing by MLE, but looking at the steps makes what is going on clearer (at least for me)

Step 1: Estimation of Choice Model

The probability of choosing $J_i = f$ is:

$$\begin{aligned}
\Pr(\, J_i = f \mid X_i = x) &= \Pr(Y_{fi} > Y_{hi} \mid X_i = x) \\
&= \Pr\left(x_0'\gamma_{0f} + x_f'\gamma_{ff} + \varepsilon_{fi} > x_0'\gamma_{0h} + x_h'\gamma_{hh} + \varepsilon_{hi}\right) \\
&= \Pr\left(x_0'\left(\gamma_{0f} - \gamma_{0h}\right) + x_f'\gamma_{ff} - x_h'\gamma_{hh} > \varepsilon_{hi} - \varepsilon_{fi}\right) \\
&= \Phi\left(\frac{x_0'\left(\gamma_{0f} - \gamma_{0h}\right) + x_f'\gamma_{ff} - x_h'\gamma_{hh}}{\sigma^*}\right) \\
&= \Phi\left(x'\gamma^*\right)
\end{aligned}$$

where Φ is the cdf of a standard normal, σ^* is the standard deviation of $(\varepsilon_{hi} - \varepsilon_{fi})$ (recall that if $\varepsilon_{fi}, \varepsilon_{hi}$ normal, then $(\varepsilon_{hi} - \varepsilon_{fi})$ normal) and

$$\gamma^* \equiv \left(\frac{\gamma_{0f} - \gamma_{0h}}{\sigma^*}, \frac{\gamma_{ff}}{\sigma^*}, \frac{-\gamma_{hh}}{\sigma^*}\right).$$

From the choice model alone we can only identify γ^*

This is referred to as the “reduced form probit”

can be estimated by maximum likelihood as a probit model

Let $\widehat{\gamma^*}$ represent the estimated parameter.

Step 2: Estimating the Wage Equation

This is essentially the second stage of a Heckman two step. To review the idea behind that, let

$$\varepsilon_i^* = \frac{\varepsilon_{hi} - \varepsilon_{fi}}{\sigma^*}$$

Then consider the regression

$$\varepsilon_{fi} = \tau \varepsilon_i^* + \zeta_i$$

where $cov\left(\varepsilon_i^*, \zeta_i\right) = 0$ (by definition of regression) and thus:

$$\begin{aligned}
\tau &= \frac{cov\left(\varepsilon_{fi}, \varepsilon_i^*\right)}{var\left(\varepsilon_i^*\right)} \\
&= E\left[\varepsilon_{fi}\left(\frac{\varepsilon_{fi} - \varepsilon_{fi}}{\sigma^*}\right)\right] \\
&= \frac{\sigma_f^2 - \sigma_{fh}}{\sigma^*}
\end{aligned}$$

Now notice that

$$\begin{aligned} E\left(Y_i \mid J_i = f, X_i = x\right) &= x_0'\gamma_{0f} + x_f'\gamma_{ff} + E\left(\varepsilon_{fi} \mid J_i = f, X_i = x\right) \\ &= x_0'\gamma_{0f} + x_f'\gamma_{ff} + E\left(\tau\varepsilon_{fi}^* + \zeta_i \mid \varepsilon_i^* > x'\gamma^*\right) \\ &= x_0'\gamma_{0f} + x_f'\gamma_{ff} + \tau E\left(\varepsilon_{fi}^* \mid \varepsilon_i^* > x'\gamma^*\right) \\ &= x_0'\gamma_{0f} + x_f'\gamma_{ff} + \tau\lambda\left(x'\gamma^*\right) \end{aligned}$$

where $\lambda\left(x'\gamma^*\right) = \frac{\phi(x'\gamma^*)}{(1-\Phi(x'\gamma^*))}$ is the inverse Mills ratio.

OLS of Y_i on X_{0i}, X_{fi}, and $\lambda\left(X_i'\widehat{\gamma^*}\right)$ gives consistent estimates of γ_{0f}, γ_{ff}, and τ

Since λ is a nonlinear function we don't have to have an exclusion restriction

Step 3: The Structural Probit

Our next goal is to estimate γ_{0h} and γ_{hh}. Note that at this point we have shown how to obtain consistent estimates of

$$\gamma^* \equiv \left(\frac{\gamma_{0f} - \gamma_{0h}}{\sigma^*}, \frac{\gamma_{ff}}{\sigma^*}, \frac{-\gamma_{hh}}{\sigma^*} \right)$$

But from the Heckman Two step we got a consistent estimates of γ_{0f} and γ_{ff}

Thus analogous to Method 1 above, as long as we have an exclusion restriction X_{fi} we can identify σ^*

Once we have σ^* it is easy to see how to identify γ_{hh} and γ_{0h}

In terms of estimation of these objects the typical way is like the second step described above.

We can estimate the “structural probit:”

$$Pr(J_i = f \mid X_i = x) = \Phi\left(\frac{1}{\sigma^*}\left(x_0'\gamma_{0f} + x_f'\gamma_{ff}\right) - x_0'\frac{\gamma_{0h}}{\sigma^*} - x_h'\frac{\gamma_{hh}}{\sigma^*}\right). \tag{1}$$

That is one just runs a probit of J_i on $\left(X_{0i}'\widehat{\gamma_{0f}} + X_{fi}'\widehat{\gamma_{ff}}\right)$, X_{0i}, and X_{hi}.

Again for identification we need an X_{fi}

Step 4: Identification of the Variance Covariance Matrix of the Residuals

Lastly, we identify all the components of Σ, $(\sigma_f^2, \sigma_h^2, \sigma_{fh})$ as follows. First, we have identified $(\sigma^*)^2 = \sigma_f^2 + \sigma_h^2 - \sigma_{fh}^2$. Second, we have identified $\tau = \frac{\sigma_f^2 - \sigma_{fh}}{\sigma^*}$. This gives us two equations in three parameters. We can obtain the final equation by using the variance of the residual in the selection model since as Heckman and Honore point out

$$\sigma_f^2 = Var(Y_i \mid J_i = f, X_i = x) - \tau^2 \left(\lambda(x'\gamma^*) x'\gamma^* - \lambda^2(x'\gamma^*) \right)$$

$$\sigma_{fh} = \sigma_f^2 - \tau\sigma^*$$

$$\sigma_h^2 = \sigma^{*2} - \sigma_f^2 + 2\sigma_{fh}.$$

Outline

Basic Model

Implications of Model

Normal Random Variables

Identification

Parametric Roy Model

Nonparametric Identification of Roy Model

Nonparametric Identification

Next we consider nonparametric identification of the Roy model

We consider the model

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

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

The formal assumptions can be found in Heckman and Honore or my chapter with French, but let me mention two

I get some normalizations, I am going to normalize the medians of ε_{fi} and $\varepsilon_{fi} - \varepsilon_{hi}$ to zero

The really strong assumption is $supp(g_f(X_{fi}, x_0), g_h(X_{hi}, x_0)) = \mathbb{R}^2$ for all $x_0 \in supp(X_{0i})$.

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

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

where G_{h-f} is the distribution function for $\varepsilon_{ih} - \varepsilon_{if}$ and $g^*(x) \equiv g_f(x_f, x_0) - g_h(x_h, x_0)$.

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

$$g^*(X_i) \geq \varepsilon_{ih} - \varepsilon_{if}$$

if and only if

$$M(g^*(X_i)) \geq M(\varepsilon_{ih} - \varepsilon_{if}).$$

A very convenient normalization is to choose the uniform distribution for $\varepsilon_{ih} - \varepsilon_{if}$.

Note that for any random variable ε with cdf F,

$$\begin{aligned} F(x) &\equiv Pr(\varepsilon \leq x) \\ &= Pr(F(\varepsilon) \leq F(x)) \end{aligned}$$

Thus $F(\varepsilon)$ has a uniform distribution.

This is a really nice normalization, we let M be G_{h-r} and define

$$\begin{aligned}\widehat{\varepsilon}_i &= G_{h-r}(\varepsilon_{ih} - \varepsilon_{if}) \\ \widehat{g}^*(x) &= G_{h-r}(g^*(x))\end{aligned}$$

Then

$$\begin{aligned}Pr(J_i = f \mid X_i - x) &= Pr(\varepsilon_{ih} - \varepsilon_{if} < g^*(x)) \\ &= Pr(G_{h-r}(\varepsilon_{ih} - \varepsilon_{if}) < G_{h-r} g^*(x)) \\ &= Pr(\widehat{\varepsilon}_i < \widehat{g}^*(x)) \\ &= \widehat{g}^*(x).\end{aligned}$$

Thus we have thus established that we can write the model as $J_i = f$ if and only if $\widehat{g}^*(X_i) > \widehat{\varepsilon}_i$ 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 $(X_i = x, J_i = f.)$

In particular we can identify

$$\begin{aligned} Med(Y_i \mid X_i = x, J_i = f) =& g_f(x_f, x_0) \\ & + Med(\varepsilon_{fi} \mid \widehat{\varepsilon}_i < \widehat{g}^*(x)). \end{aligned}$$

An exclusion restriction is key.

For any (x_f^a, x_0^a) and (x_f^b, x_0^b) if I can find values of x_h^a and x_h^b such that

$$\widehat{g}^*(x^a) = \widehat{g}^*(x^b)$$

where $x^a = (x_0^a x_f^a, x_h^a)$ and $x^b = (x_0^b x_f^b, x_h^b)$

then

$$\begin{aligned}
&Med(Y_i \mid X_i = x^a, J_i = f) - Med(Y_i \mid X_i = x^b, J_i = f)\\
&\quad = g_f(x_f^a, x_0^a) - g_f(x_f^a, x_0^a)\\
&\qquad + Med(\varepsilon_{fi} \mid \widehat{\varepsilon}_i < \widehat{g}^*(x^a)) - Med(\varepsilon_{fi} \mid \widehat{\varepsilon}_i < \widehat{g}^*(x^b))\\
&\quad = g_f(x_f^a, x_0^a) - g_f(x_f^a, x_0^a)
\end{aligned}$$

Identification at Infinity

What about the location?

Notice that

$$\begin{aligned}
&\lim_{\widehat{g}^*(x)\to 1} Med(Y_{fi} \mid X_i = x, J = f) \\
&\qquad = g_f(x_f, x_0) + \lim_{\widehat{g}^*(x)\to 1} Med(\varepsilon_{fi} \mid \widehat{\varepsilon}_i < \widehat{g}^*(x)) \\
&\qquad = g_f(x_f, x_0) + Med(\varepsilon_{fi} \mid \widehat{\varepsilon} < 1) \\
&\qquad = g_f(x_f, x_0) + Med(\varepsilon_{fi}) \\
&\qquad = g_f(x_f, x_0).
\end{aligned}$$

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 the distribution of Y_{fi} conditional on $\widehat{\varepsilon}_i > g^u$ 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.

Whether it is a big deal or not depends on the question of interest.

Step 3: Identification of g_h

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

Recall that

$Pr(J_i = f \mid X_i = (x_f, x_r, x_0)) = \Pr\left(\varepsilon_{hi} - \varepsilon_{fi} \leq g_f(x_f, x_0) - g_h(x_h, x_0)\right)$

But note that this is a cdf and that the median of $\varepsilon_{hi} - \varepsilon_{fi}$ is 0

This means that when

$$Pr(J_i = f \mid X_i = (x_f, x_r, x_0)) = 0.5,$$

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

Since g_f is identified, clearly g_h 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

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

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

Thus we know the joint distribution of $(\varepsilon_{hi} - \varepsilon_{fi}, \varepsilon_{fi})$

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