The Roy Model

Christopher Taber

January 19, 2009

Outline

Basic Model

Implications of Model

Normal Random Variables

Heckman and Honore

Outline

Basic Model

Implications of Model

Normal Random Variables

Heckman and Honore

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

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

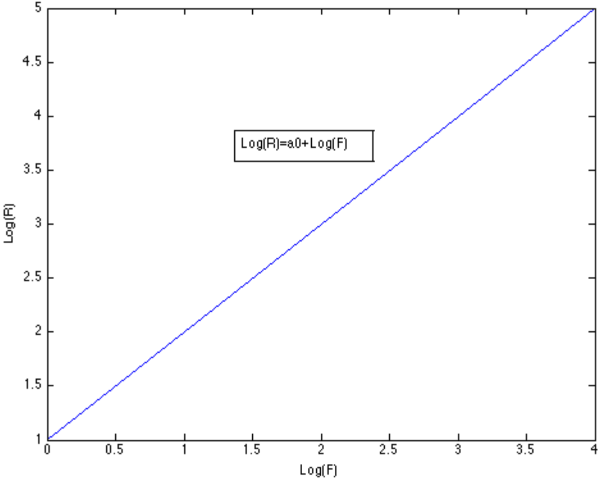
$$\log(\pi_R) + \log(R) = \log(\pi_F) + \log(F)$$

which can be written as

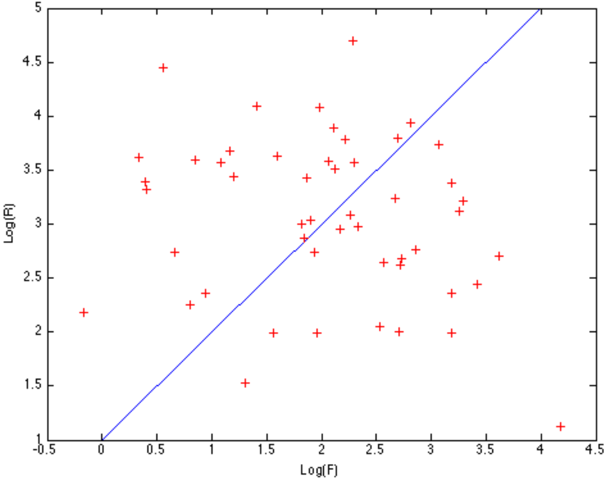
$$\log(R) = \log(\pi_F) - \log(\pi_R) + \log(F)$$

If you are above this line you hunt

If you are below it you fish



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

Case 1: No variance in Rabbits

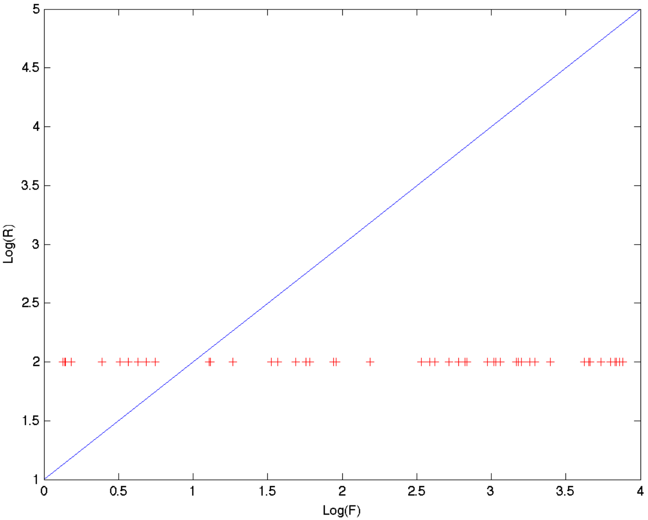
Suppose everyone catches R^*

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

Case 2: Perfect correlation

Suppose that

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

with $\alpha_1 > 0$

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

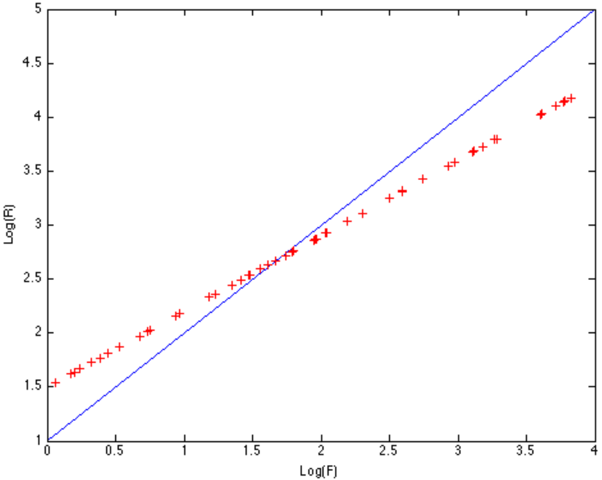
Fish if

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

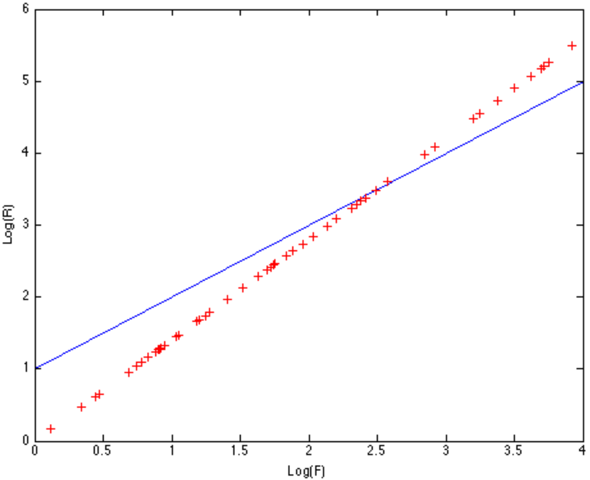
If $\alpha_1 < 1$ then left hand side is increaing in $\log(F)$ meaning that better fisherman are more likely to fish

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)

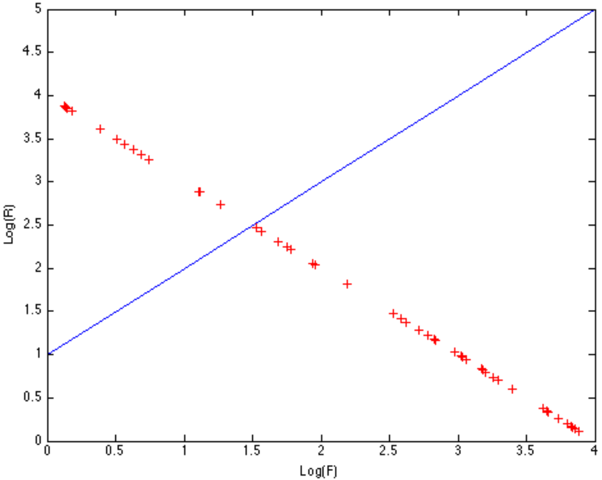


Log(R)
Log(F)
0
0.5
1
1.5
2
2.5
3
3.5
4
5
6

Case 3: Perfect Negative Correlation

Exactly as before

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



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

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

Heckman and Honore

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

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

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

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

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

Then

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

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

$$u_F = \alpha\,(u_F - u_R) + \omega$$

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

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

The question boils down to the sign of this object.

If it is positive then positive selection into fishing

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

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

Notice that

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

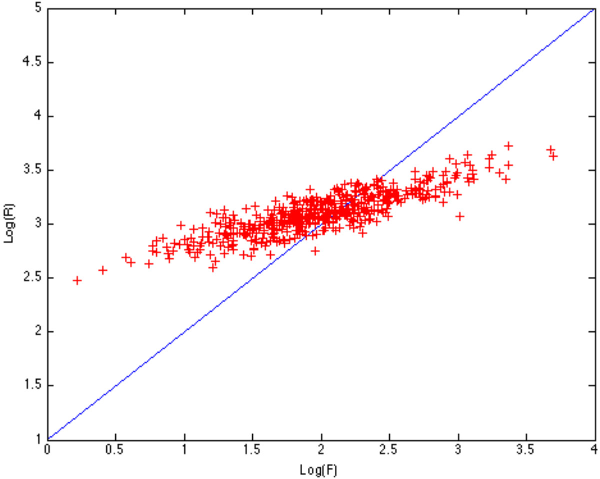
One of these must be positive.

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

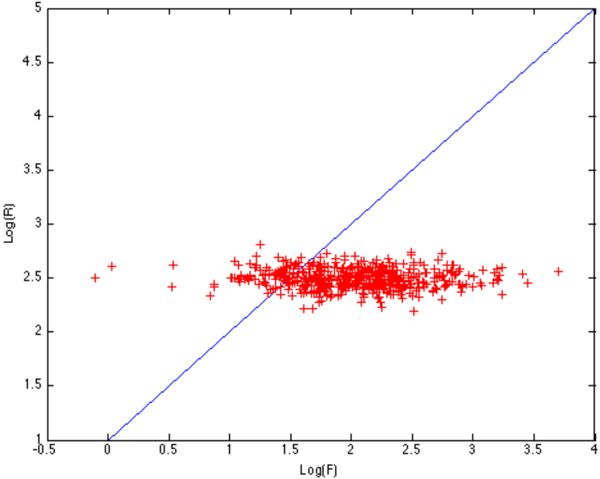
Hunters could go either way depending on $\sigma_{RR} - \sigma_{FR}$

- 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)



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

Outline

How do we think about estimating this model?

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

Lets think about how we would estimate this 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

We can identify

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

If R and F are lognormal this is enough

However, if we don't know their distribution it isn't

For example independent skills can explain data perfectly (Heckman and Honore)

To identify this we need more data

Lets suppose we have multiple periods or multiple villages

Furthermore suppose that prices are known

To keep thinks simple lets condition on $\pi_F = 1$

However we will let π_R vary from $(0, \infty)$

What can we identify?

$$
\begin{aligned}
\Pr(W \leq x; \pi_R) &= \Pr(\max\{F, \pi_R R\} \leq x; \pi_R) \\
&= \Pr(F \leq x, \pi_R R \leq x; \pi_R) \\
&= \Pr\left(F \leq x, R \leq \frac{x}{\pi_R}; \pi_R\right) \\
&= G\left(x, \frac{x}{\pi_R}\right)
\end{aligned}
$$

By moving x and π_R around we can identify G

Depended on 4 assumptions:

- Roy model (only wages matter)
- Full support of prices
- *G* stable across time (villages)
- Prices known

The last assumption can be relaxed in a number of ways, for example panel data

For example we observe a person at two points in time who fishes in both periods

$$\begin{aligned}\frac{W_{it}}{W_{i\tau}} &= \frac{\pi_{Ft}F_i}{\pi_{F\tau}F_i} \\ &= \frac{\pi_{Ft}}{\pi_{F\tau}}\end{aligned}$$

since there is always some normalization, ratios are enough

Variation in observables also helps

Suppose

$$\begin{aligned} \log(F) &= g_F(X_F, X_0) + u_F \\ \log(R) &= g_R(X_R, X_0) + u_R \end{aligned}$$

where (u_F, u_R) is independent of (X_F, X_R, X_0)

X_F and X_R are both exclusion restrictions

Now normalize $\pi_R = \pi_F = 1$

Fish if

$$u_F - u_R > g_R(X_R, X_0) - g_F(X_F, X_0)$$

Then

$$\begin{aligned}
&\Pr(\text{fish}, \ \log(F) < x) \\
&= \Pr(u_R - u_F \leq g_F(X_F, X_0) - g_R(X_R, X_0), u_F < x - g_F(X_F, X_0)) \\
&= G_{u_F - u_R, u_F}(g_F(X_F, X_0) - g_R(X_R, X_0), x - g_F(X_F, X_0))
\end{aligned}$$

From this you can get the joint distribution of u_R and u_F.