Notes on Roy Models and Generalized Roy (Extract)

James J. Heckman
University of Chicago

Econ 312, Spring 2022

Basic Framework of Roy Model

- Agents possess advantages in tasks associated with sector j, $j \in \mathcal{J}$.
- They get an income Y_j for participating in sector j. ($Y_{j,i}$ for agent i)
- There may be a cost C_j of participating in the sector ($C_{j,i}$ for person i).
- A one-period model (will extend to multiple periods later)
- When making their choices, they are uncertain and have information set $\mathcal{I}_i$.

- For notational simplicity, drop the i subscript.
- Agents select sector $\hat{j}$ such that

$$\hat{j} = \underset{j \in \mathcal{J}}{\operatorname{argmax}} E\left\{\{Y_j - C_j\} \mid \mathcal{I}\right\}$$

- Toss a coin in the event of a tie.
- Ties are often assumed away as negligible events (i.e., absolute continuity is assumed).

- Ex post agents may regret their choices. E.g.,

$$Y_{\hat{j}} - C_{\hat{j}} < 0$$

or even

$$\left(Y_{\hat{j}} - C_{\hat{j}}\right) < \left(Y_j - C_j\right)_{j \in \mathcal{J} \setminus \{\hat{j}\}}$$

- The Y_j can be a variety of outcomes.
  Examples:
  1. Different labor force states (work, not work) and C_j is cost of working

     $$e.g., Y_1 = \text{ value of market time}$$
     $$Y_0 = \text{ value of home production}$$

     so if $C_1 = 0$ and $C_0 = 0$
     Y_1 is the market wage
     Y_0 is the reservation wage
     Reservation wage can come from
     1. Search Theory (see, e.g., Shimer, 2010)
     2. Value of Time in the home (see, e.g., Heckman, 1974; Mulligan and Rubinstein, 2008)
  2. Earnings in different countries (Borjas, 1987)

3. Earnings in different occupations (Miller, 1984; Jovanovic, 1979a,b; Pavan, 2008)
4. Earnings at different schooling levels (e.g., Willis and Rosen, 1979; Keane and Wolpin, 1997, 2011; Heckman, Lochner, and Taber, 1998; Johnson, 2013; Heckman, Humphries and Veramendi, 2018.)
5. Randomization bias (Kline and Walters, 2016)

- Under the earnings interpretation, let π_j be the price of skill j (the rental rate or the return)
- The quantity of skill j is S_j
- $Y_j = \pi_j' S_j$ (gross earnings)
- $Y_j - C_j = \pi_j' S_j - C_j$ (net earnings)

The Roy Model: Example

Two sector Roy model. (sectors $j \in \{1, 2\}$)

Income maximizing agents possess two skills $S_1 = s_1$ and $S_2 = s_2$ with associated positive skill prices π_1 and π_2.

Skills are scalar (for now)

Agent chooses sector 1 if his earnings are greater there

$$W_1 = Y_0 = \pi_1 S_1$$
$$W_1 = Y_1 = \pi_1 S_1$$
$$\pi_1 S_1 > \pi_2 S_2$$

Proportion of the population working in sector one,

$$P_1 = \Pr(\pi_1 S_1 > \pi_2 S_2) :$$

$$P_1 = \int_0^{\infty} \int_0^{\pi_1 s_1 / \pi_2} f(s_1, s_2) ds_2 ds_1 \tag{2.1}$$

Density of skill employed in sector one differs from the population density of skill. (selection problem)

The latter density:

$$f_1(s_1) = \int_0^{\infty} f(s_1, s_2) ds_2.$$

Former density:

$$g(s_1 \,|\pi_1 s_1 > \pi_2 s_2) = \frac{1}{P_1} \int_0^{\pi_1 s_1/\pi_2} f(s_1, s_2) ds_2$$

Density of earnings in sector 1 (using $w_1 = \pi_1 s_1$):

$$g_1(w_1) = \frac{1}{P_1 \pi_1} \int_0^{w_1/\pi_2} f(w_1/\pi_1, s_2) ds_2$$

Similarly, the density of skill employed in sector 2 is:

$$g(s_2 \,|\pi_2 s_2 > \pi_1 s_1) = \frac{1}{P_2} \int_0^{\pi_2 s_2/\pi_1} f(s_1, s_2) ds_1$$

The density of earnings in sector two is:

$$g_2(w_2) = \frac{1}{P_2 \pi_2} \int^{w_2/\pi_1} f(s_1, \frac{w_2}{\pi_2}) ds_1 \tag{2.2}$$

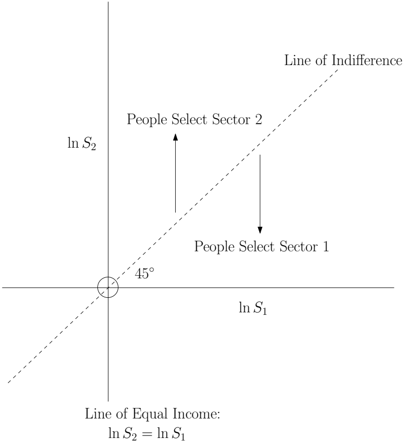
The overall density of earnings is:

$$g(w) = P_1 g_1(w) + P_2 g_2(w)$$

(A mixture of two densities)

Set $\pi_1 = \pi_2 = 1$

Take logs: Partitions of $(\ln S_2, \ln S_1)$ space:



- As $\pi_2 \uparrow$, line shifts down in parallel fashion.

Normal Roy Model: Some Illustrations

$$(\ln S_1, \ln S_2) \sim N(\mu_1, \mu_2, \Sigma)$$

$$\begin{aligned} E(\ln S_j) &= \mu_j \\ \ln S_j &= \mu_j + U_j \\ \Rightarrow \ln W_j &= \ln \pi_j + \mu_j + U_j, \; j = 1, 2 \end{aligned} \tag{3.1}$$

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right)$$

Define

$$\sigma^* = [Var(U_1 - U_2)]^{1/2} = \sqrt{\sigma_{11} + \sigma_{22} - 2\sigma_{12}}$$
$$c_1 = (\ln(\frac{\pi_1}{\pi_2}) + \mu_1 - \mu_2)/\sigma^*.$$

Define:

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{\frac{-q2}{2}} dq$$
$$P_1 = P(\ln W_1 > \ln W_2) = 1 - \Phi(-c_1) = \Phi(c_1)$$

Choice equation (can be of very general functional form)

Line of Indifference:

$$\ln W_1 - \ln W_2 = \ln\left(\frac{\pi_1}{\pi_2}\right) + \mu_1 - \mu_2 + U_1 - U_2$$
$$L = U_1 - U_2$$
$$c_1^* = \ln(\pi_1/\pi_2) + \mu_1 - \mu_2.$$

$$\begin{aligned} & E\left(\ln W_1 \mid \ln W_1 - \ln W_2 > 0\right) \\ & \quad = \ln \pi_1 + \mu_1 + \underbrace{E(U_1 \mid L > -c_1^*)}_{\text{Selection Bias Term}}. \end{aligned} \tag{3.2}$$

(For estimation: Control Function)

Selection operates through the dependence between U_1 and $(U_1 - U_2)$.

More generally through the unobservables in the $\ln W_1$ and the decision equation. $(I = Y_2 - Y_1 - (C_2 - C_1))$

Observe Y_2 if $Y_1 - Y_1 - (C_2 - C_1) > 0$
(Censoring condition and Y_2 is a censored random variable)

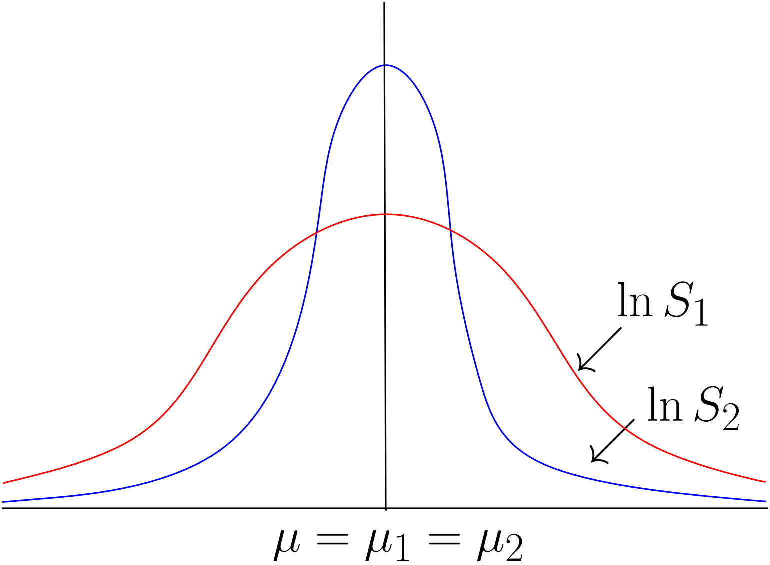
Observe Y_1 otherwise

Can We Get Negative Selection in Sector 2?

Arises if $\sigma_{22} < \sigma_{12} < \sigma_{11}$

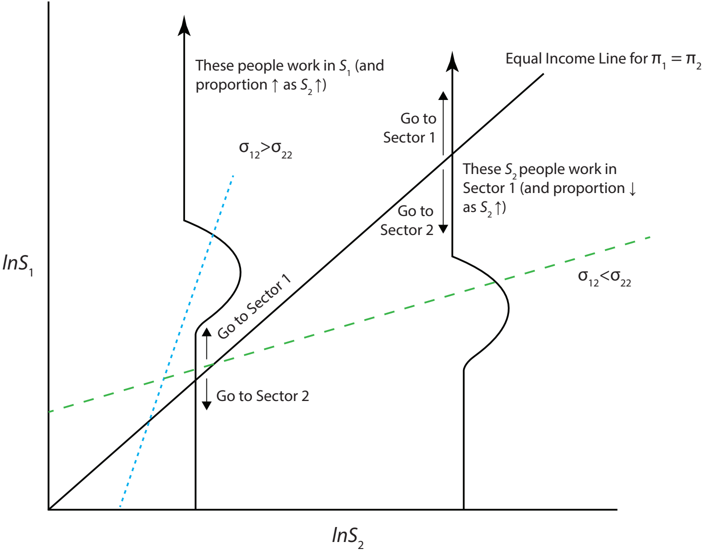
Example: Set $\pi_1 = \pi_2$

- $D = \mathbf{1}\left(\ln S_1 > \ln S_2\right)$
- $\sigma_{22} \leq \sigma_{12} \leq \sigma_{11}$
- $\mu_1 = \mu_2$



Densities of $\ln S_1$ and $\ln S_2$

- People selected into S_2 are below average in 2.
- People selected in S_1 are above average in 1.



$$lnS_1 = \mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(lnS_2 - \mu_2) + \upsilon_1$$

- Notice: Axes switched from previous figure.