V. J. Hotz
Latest Revision: May 16, 2007

The Roy Model of Self-Selection: Simple Case

A core topic in labor economics is 'self-selection.' What this term means in theory is that rational actors make optimizing decisions about what markets to participate in — job, location, education, marriage, crime, etc. What it means in practice is that observed economic relationships should generally be viewed as endogenous outcomes of numerous optimizing decisions, rather than as exogenous causal relationships. Understanding self-selection should make you skeptical of treating any ecological correlation as causal.

The starting point of formal treatment of this topic in economics is Roy's (1951) "Thoughts on the Distribution of Earnings," which discusses the optimizing choices of 'workers' selecting between fishing and hunting. Roy's key observation is that there are three factors that affect this choice:

1. Fundamental distribution of skills and abilities

2. The correlations among these skills in the population

3. The technologies for applying these skills

4. Consumer tastes that impact demand for different types of outputs

At the time of Roy's writing, the presumption was that the distribution of income that arises from economic processes is arbitrary. Hence, if we compare the mean earnings of hunters and fishermen, $\bar{y}_h$ and $\bar{y}_f$, then $\bar{y}_h - \bar{y}_f$ is an estimate of the earnings gain or loss that an individual would receive from switching from fishing to hunting. Roy's article explains why this view is incorrect.

The essential departure of Roy's model from previous work is that it is a multiple-index model (in this case, 2 indices): workers have skills in each occupation, but they can only use one skill or the other. Hence, workers self-select the sector that gives them the highest expected earnings. Equilibrium in each market equates supply and demand, while a self-selection condition means that the marginal worker is indifferent between the two sectors.

In these notes, I lay out the simpliest version of the Roy Model. To do so, I follow the structure of the paper by George Borjas, "Self-Selection and the Earnings of Immigrants" in the *AER*, 1987. This paper characterizes a simple, parametric 2-sector Roy model. The enduring contribution of Borjas' paper is this model rather than the empirical findings. Any labor economist should be well versed with this model.

1. Borjas 1987: Self-Selection and the Earnings of Immigrants

Who chooses to immigrate to the United States? One ready-made answer is that workers from low wage countries will immigrate. This may be true on average, but it's probably too simple. The workers immigrating to the United States are probably not a random subset of the Mexican workforce. Rather, we should expect that potential migrants make some rough comparison of their wages in the home country and their expected wages in the U.S. On average, we'd expect those who immigrate to have higher expected earnings in the U.S. than Mexico and vice versa for those who stay.

- Consider two countries Mexico (0) and the United States (1), denoting the "source" and "host" country, respectively.

- Log earnings in the source country are given by

$$w_0 = \mu_0 + \varepsilon_0 \tag{1}$$

  where $\varepsilon_0 \sim N(0, \sigma_0^2)$. It's useful to think of ε_0 as the de-meaned value of worker's 'skill' in Mexico (the source country).

- If everyone from Mexico were to migrate to the U.S. (host country), their earnings would be (ignoring any general equilibrium effects!):

$$w_1 = \mu_1 + \varepsilon_1 \tag{2}$$

  where $\varepsilon_1 \sim N(0, \sigma_1^2)$.

- Assume that the cost of migrating is C, which Borjas puts into 'time equivalent' terms as $\pi = C/w_0$. Borjas further assumes that π is constant, meaning that C is directly proportional to w_0.

- Assume further that each worker knows C, μ_0, μ_1 and his individual epsilons: ε_0, ε_1.

- You, the econometrician, only observe a worker in one country or the other, and hence you only know ε_0 or ε_1 for any individual.

KEY QUESTIONS:

1. What can you infer about what wages for Mexican immigrants in the United States would have been had they stayed in their source countries?

2. What would wages in the United States be for non-migrant Mexicans had they come to the U.S.?

The Roy Model answers these questions.

- The *correlation* between the unobservable (ε) component of source (0) and host (1) country earnings is

$$\rho = \frac{\sigma_{01}}{\sigma_0 \sigma_1} \tag{3}$$

where $\sigma_{01} = \text{cov}(\varepsilon_0, \varepsilon_1)$.

- To implement this model, we need to know ρ, although we do not need to know both ε_0, ε_1 for any worker.

Self-Selection Decision Rule: *A Mexican worker will choose to migrate to the U.S. iff*

$$(\mu_1 - \mu_0 - \pi) + (\varepsilon_1 - \varepsilon_0) > 0. \tag{4}$$

Define the indicator variable $I = 1$, *if this selection condition is satisfied and* $I = 0$, *otherwise*.

♦ Now, define $v \equiv \varepsilon_1 - \varepsilon_0$. The probability that a randomly chosen worker from Mexico (the source country) chooses to migrate to the U.S. (the host country) is equal to

$$\begin{aligned} P &= \Pr\left[I = 1\right] \\ &= \Pr\left[v > (\mu_0 - \mu_1 + \pi)\right] \\ &= \Pr\left[\frac{v}{\sigma_v} > \frac{(\mu_0 - \mu_1 + \pi)}{\sigma_v}\right] \\ &= 1 - \Phi\left(\frac{(\mu_0 - \mu_1 + \pi)}{\sigma_v}\right) \\ &= 1 - \Phi(z) \end{aligned} \tag{5}$$

where $z = \frac{(\mu_0 - \mu_1 + \pi)}{\sigma_v}$ and $\Phi(\cdot)$ is the CDF of the standard normal. Notice that the higher larger is z, the lower is the probability of migration (from Mexico to the U.S). This is because z is rising in the mean earnings of Mexico (μ_0) and the cost of migration (π). So it follows that

$$\frac{\partial P}{\partial \mu_0} < 0, \frac{\partial P}{\partial \mu_1} > 0, \frac{\partial P}{\partial \pi} < 0.$$

♦ These are the ***mean effects*** and drive, in part, self-selection of agents. In our example, high mean wages in the U.S. (μ_1) relative to those in Mexico (μ_0), create a net incentive for workers to migrate from Mexico to the U.S. Variation in the "other" costs of migration also create such incentives.

♦ However, there is more to the nature of self-selection in the Roy Model. To see these implications of this model, it is useful to assume from here forward that

$$\mu_0 \approx \mu_1, \tag{6}$$

so we can focus on self-selection properties of the simple Roy Model rather than its implications for mean differences.

1.1 Selection Conditions

What is the expectation of earnings in the source country (Mexico) for workers who choose to migrate to the U.S.?

$$\begin{aligned} E\left(w_0 \middle| I=1\right) &= \mu_0 + E\left(\varepsilon_0 \middle| \frac{\nu}{\sigma_\nu} > z\right) \\ &= \mu_0 + \sigma_0 E\left(\frac{\varepsilon_0}{\sigma_0} \middle| \frac{\nu}{\sigma_\nu} > z\right) \end{aligned} \tag{7}$$

- Notice that this equation depends on three things:

  1. Mean earnings in the source country (μ_0),

  2. Both error terms (ε_1, ε_0) through ν,

  3. Implicitly, it also depends on the *correlation* between the error terms.

- We want to know the expectation of ε_0 given some value ν. Given the normality of ε_1, ε_0, this is simply equal to the coefficient on the following *conditional expectation function*:

$$E(\varepsilon_0 | \nu) = \frac{\sigma_{0\nu}}{\sigma_\nu^2}\nu. \tag{8}$$

  where $\sigma_{0\nu} \equiv \text{cov}(\varepsilon_0, \nu) = \sigma_{01} - \sigma_0^2$. Note that the conditional expectation in (8) implies that we can always write following *population regression function*:

$$\varepsilon_0 = \frac{\sigma_{0\nu}}{\sigma_\nu^2}\nu + \xi \tag{9}$$

  where ξ is a mean zero error term that is independent of ν. (There is no intercept in either (8) or (9), given that ε_0 and ν both have zero means.)

Using (8), it follows that:

$$E\left(\frac{\varepsilon_0}{\sigma_0}\middle|\frac{\nu}{\sigma_\nu}\right)=\frac{\sigma_{0\nu}}{\sigma_\nu^2}\cdot\sigma_\nu^2\cdot\frac{1}{\sigma_0\sigma_\nu}\cdot\frac{\nu}{\sigma_\nu}$$
$$=\frac{\sigma_{0\nu}}{\sigma_0\sigma_\nu}\frac{\nu}{\sigma_\nu} \qquad (10)$$
$$=\rho_{0\nu}\frac{\nu}{\sigma_\nu}$$

where $\rho_{0\nu}\equiv\frac{\operatorname{cov}(\varepsilon_0,\nu)}{\sigma_0\sigma_\nu}=\frac{\sigma_{01}-\sigma_0^2}{\sigma_0\sigma_\nu}$.

- It follows that we can rewrite (7) – the expected earnings that would prevail in Mexico (w_0) among those Mexicans who would choose to migrate to the U.S. – as:

$$E(w_0|I=1)=\mu_0+\sigma_0 E\left(\frac{\varepsilon_0}{\sigma_0}\middle|\frac{\nu}{\sigma_\nu}>z\right)$$
$$=\mu_0+\rho_{0\nu}\sigma_0 E\left(\frac{\nu}{\sigma_\nu}\middle|\frac{\nu}{\sigma_\nu}>z\right)$$
$$=\mu_0+\rho_{0\nu}\sigma_0\left(\frac{\phi(z)}{1-\Phi(z)}\right) \qquad (11)$$
$$=\mu_0+\rho_{0\nu}\sigma_0\left(\frac{\phi(z)}{\Phi(-z)}\right)$$
$$=\mu_0+\rho_{0\nu}\sigma_0\lambda(z)$$

where $\lambda(z)\equiv\frac{\phi(z)}{1-\Phi(z)}$ is the *Inverse Mills Ratio*, which is equal to the conditional expectation of a standard normal random variable truncated from the left at a point z. [See Heckman, "Selection Bias as a Specification Error," *Econometrica*, 1979 for more on properties of normal random variables and their expectations.]

- It follows that the expected earnings that will prevail in the U.S. for those who migrate from Mexico to the U.S. is given by:

$$
\begin{aligned}
E(w_1 | I = 1) &= \mu_1 + E\left(\varepsilon_1 \middle| \frac{\nu}{\sigma_\nu} > z \right) \\
&= \mu_1 + \rho_{1\nu}\sigma_1 \left(\frac{\phi(z)}{\Phi(-z)} \right)
\end{aligned} \tag{12}
$$

- It is convenient to rewrite (11) and (12) as follows:

$$
\begin{aligned}
E(w_0 | I = 1) &= \mu_0 + \rho_{0\nu}\sigma_0 \lambda(z) \\
&= \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\rho - \frac{\sigma_0}{\sigma_1} \right)\lambda(z)
\end{aligned} \tag{13}
$$

$$
\begin{aligned}
E(w_1 | I = 1) &= \mu_1 + \rho_{1\nu}\sigma_1 \lambda(z) \\
&= \mu_1 + \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\frac{\sigma_1}{\sigma_0} - \rho \right)\lambda(z)
\end{aligned} \tag{14}
$$

where $\rho = \dfrac{\sigma_{01}}{\sigma_0\sigma_1}$. [**You should verify the above to expressions!**]

1.2 Ways in Which Self-Selection Alter Expected Wages

- How do the mean earnings of those who self-select to migrate from Mexico to the U.S. compare to the mean earnings in Mexico?

- How do they compare to the mean earnings in the U.S.?

- There are three cases to consider.

- To characterize these cases, define:

$$
Q_0 \equiv E\left(\varepsilon_0 | I = 1\right) \text{ and } Q_1 \equiv E\left(\varepsilon_1 | I = 1\right)
$$

where Q_0 and Q_1 are the truncated means of the unobserved components of earnings in Mexico (ε_0) and earnings in the U.S. (ε_1), given migration (from Mexico to the U.S.).

1.2.1 <u>Positive</u> <u>Hierarchical</u> <u>Sorting</u>

- This is a case where migrants from Mexico to the U.S. are positively selected from the source country (Mexican) distribution and are also above the mean of the host country (U.S.) distribution, i.e.,

$$Q_0 > 0 \text{ and } Q_1 > 0. \quad (15)$$

This type of sorting will occur iff:

$$\frac{\sigma_1}{\sigma_0} > 1 \text{ and } \rho > \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right). \quad (16)$$

- What do these conditions mean?

1. $\frac{\sigma_1}{\sigma_0} > 1$ implies that the host country (U.S.) has a more "dispersed" income than does the country of origin (Mexico). In effect, this condition implies that the U.S. has a higher "return to skill" than does Mexico.

2. $\rho > \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right)$ implies that the correlation between the skills valued in the host (U.S.) and source (Mexico) countries is "sufficiently high." If you were a skilled worker in Mexico, you would not want to migrate to the U.S. a host country with a very high return to skills if the skills valued in the U.S. were uncorrelated (or negatively correlated) with skills value in Mexico (the country of origin).

- This case embodies the canonical American view of immigration: 'The best and the brightest' leave their home countries for greater opportunity (that is, higher return to skill) in the U.S.

- One way of restating this type of migration is: a source country with low earnings variance 'taxes' the earnings of high skill workers and insures the earnings of low skill workers. High skill workers may want to emigrate, accordingly. But this is not the only possibility.

1.2.2 Negative Hierarchical Sorting

- This is a case where migrants from Mexico (country of origin) to the U.S. (host country) are *negatively self-selected* from the source country distribution and are also below the average of the host country distribution. That is:

$$Q_0 < 0 \text{ and } Q_1 < 0. \tag{17}$$

This type of sorting will occur iff:

$$\frac{\sigma_1}{\sigma_0} < 1 \text{ and } \rho > \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right). \tag{18}$$

- This is simply the converse case. Here, the source country(Mexico) is unattractive to low earnings workers because it has high wage dispersion. Again assuming that wages are sufficiently correlated between the source and host country, low skill workers will want to migrate to the U.S. (host country) to take advantage of the 'insurance' provided by a narrower wage structure in the host country (U.S.).

- This case has the potentially unattractive character – certainly from Borjas' perspective – where a compressed wage structure 'subsidizes' low skill workers, thus attracting low skill workers from abroad.

1.2.3 'Refugee' Sorting

- A third case is where

$$Q_0 < 0 \text{ and } Q_1 > 0. \tag{19}$$

that is, migrants from Mexico to the U.S. are selected from the lower tail of the home country (Mexico) distribution but arrive in the upper tail of the host country (U.S.) distribution. This type of sorting will occur iff:

$$\rho < \min\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right). \tag{20}$$

meaning that the correlation between earnings in the two countries is sufficiently low and could even be negative, i.e., the skills valued in one country receive a low return in the other country.

- This might occur, for example, for a minority group whose opportunities in the host country are depressed by prejudice. Or for the case of migration from a non-market economy where the set of skills rewarded is quite different from the economy in the receiving country (e.g., European Jews in the first case, intellectuals from the former Communist block in the second).

1.2.4 A Fourth Case?

- Note that there is not a fourth case where $Q_0 > 0$, $Q_1 < 0$. Why not? This would suggest irrational migration, where people leave the upper tail of the source country income distribution to join the lower tail of the host country distribution. This is inconsistent with income maximization, at least in situations where $\mu_0 \approx \mu_1$.

- Mathematically, this case would require that

$$\rho > \max\left(\frac{\sigma_1}{\sigma_0}, \frac{\sigma_0}{\sigma_1}\right) \tag{21}$$

which would imply that $\rho > 1$, which cannot be true for a correlation coefficient.

1.3 Effects of Shifts in μ_i, π and σ_i Self-Selection in Earnings

To come.