

# Lecture Note: Self-Selection – The Roy Model

David H. Autor  
MIT 14.661 Spring 2003

November 14, 2003

## 1 INTRODUCTION

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 ecological correlations 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.

Roy’s 1951 paper is amusing to read to a contemporary economist because it so awkwardly straddles the line between older-style narrative economics and modern mathematical economics.

Roy is clearly writing with equations and distributions in mind (probably all written out), but he writes mostly about rabbits and fish – only occasionally interjecting that ‘therefore’ earnings will be log-normal. This is very hard to follow. Borjas’ 1987 AER Paper on “Self-Selection and the Earnings of Immigrants” is the first paper that I know that writes down a simple, parametric 2-sector Roy model. The enduring contribution of Borjas’ paper is this model (sometimes called a Borjas selection model) rather than the empirical findings. As a labor economist, you should be well versed with this model.

## 2 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 0 and 1, denoting the source and host country.
- Log earnings in the source country are given by

$$w_0 = \mu_0 + \varepsilon_0,$$

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

- If everyone from country 0 were to migrate to the host country, their earnings would be (ignoring any general equilibrium effects!):

$$w_1 = \mu_1 + \varepsilon_1,$$

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  $\pi$  is constant, meaning that  $C$  is directly proportional to  $w_0$ .
- Assume further that each worker knows  $C, \mu_0, \mu_1$  and his individual epsilons:  $\varepsilon_0, \varepsilon_1$ .
- You, the econometrician, only observe a worker in one country or the other, and hence you only know  $\varepsilon_0$  or  $\varepsilon_1$  for any individual.
- What can you infer about what wages for immigrants in the United States would have been had they stayed in their source countries? What would wages in the United States be for non-migrants had they come to the United States? The Roy Model answers these questions.
- The correlation between source and host country earnings is

$$\rho = \frac{\sigma_{01}}{\sigma_0\sigma_1},$$

where  $\sigma_{01}$  is  $\text{cov}(\sigma_0, \sigma_1)$ .

- To implement this model, we need to know  $\rho$ , although we do not need to know both  $\varepsilon_0, \varepsilon_1$  for any worker.
- A worker will choose to migrate if

$$(\mu_1 - \mu_0 - \pi) + (\varepsilon_1 - \varepsilon_0) > 0, \tag{1}$$

(Borjas defines the indicator variable  $I$ , equal to 1 if this selection condition is satisfied, 0 otherwise).

- Now, define  $\nu = \varepsilon_1 - \varepsilon_0$ . The probability that a randomly chosen worker from the source country chooses to migrate is equal to

$$\begin{aligned} P &= \Pr[\nu > (\mu_0 - \mu_1 + \pi)] \\ &= \Pr\left[\frac{\nu}{\sigma_\nu} > \frac{(\mu_0 - \mu_1 + \pi)}{\sigma_\nu}\right] \\ &= 1 - \Phi\left(\frac{(\mu_0 - \mu_1 + \pi)}{\sigma_\nu}\right) \\ &\equiv 1 - \Phi(z), \end{aligned}$$

where  $z = (\mu_0 - \mu_1 + \pi) / \sigma_\nu$ , and  $\Phi(\cdot)$  is the CDF of the standard normal. Notice that the higher larger is  $z$ , the lower is the probability of migration. This is because  $z$  is rising in the mean earnings of the home country and the cost of migration. So  $\partial P / \partial \mu_0 < 0, \partial P / \partial \mu_1 > 0, \partial P / \partial \pi < 0$ .

- These are mean effects. It's useful to assume from here forward that  $\mu_1 \approx \mu_0$ , so we can focus on self-selection rather than mean differences.
- Let's examine the self-selection properties.

## 2.1 SELECTION CONDITIONS

- What is the expectation of earnings in the *source* country for workers who choose to immigrate?
- 

$$\begin{aligned} E(w_0 | \text{Immigrate}) &= \mu_0 + E\left(\varepsilon_0 \mid \frac{\nu}{\sigma_\nu} > z\right) \\ &= \mu_0 + \sigma_0 E\left(\frac{\varepsilon_0}{\sigma_0} \mid \frac{\nu}{\sigma_\nu} > z\right). \end{aligned} \tag{2}$$

- Notice that this equation depends on three things:
  1. Mean earnings in the source country
  2. Both error terms  $(\varepsilon_0, \varepsilon_1)$  through  $\nu$ .
  3. Implicitly, it also depends on the correlation between the error terms.
- We want to know the expectation of  $\varepsilon_0$  given some value  $\nu$ . Given the normality of  $\varepsilon_0, \varepsilon_1$ , this is simply equal to the regression coefficient

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

Applying this to (2), we get

$$\begin{aligned}
E\left(\frac{\varepsilon_0}{\sigma_0} \middle| \frac{\nu}{\sigma_\nu}\right) &= \frac{\sigma_{0\nu}}{\sigma_\nu^2} \cdot \frac{1}{\sigma_\nu^{-2}} \cdot \frac{1}{\sigma_0\sigma_\nu} \cdot \frac{\nu}{\sigma_\nu} \\
&= \frac{\sigma_{0\nu}}{\sigma_0\sigma_\nu} \frac{\nu}{\sigma_\nu} \\
&= \rho_{0\nu} \frac{\nu}{\sigma_\nu}.
\end{aligned}$$

Notice that due to the normalizations, the covariance  $\text{cov}(\varepsilon_0, \varepsilon_1)$  is reduced by  $1/\sigma_0\sigma_1$  and the variance of  $\nu/\sigma_\nu$  is 1.

- Hence, we can rewrite (2) as

$$\begin{aligned}
E(w_0|\text{Immigrate}) &= \mu_0 + \sigma_0 E\left(\frac{\varepsilon_0}{\sigma_0} \middle| \frac{\nu}{\sigma_\nu} > z\right) \\
&= \mu_0 + \sigma_0 E\left(\frac{\varepsilon_0}{\sigma_0} \middle| \frac{\nu}{\sigma_\nu} > z\right) \\
&= \mu_0 + \rho_{0\nu}\sigma_0 E\left(\frac{\nu}{\sigma_\nu} \middle| \frac{\nu}{\sigma_\nu} > z\right) \\
&= \mu_0 + \rho_{0\nu}\sigma_0 \left(\frac{\phi(z)}{1 - \Phi(z)}\right)
\end{aligned} \tag{3}$$

where  $\phi(z)/(1 - \Phi(z))$  is the Inverse Mills Ratio, equal to the conditional expectation of a standard normal random variable truncated from the left at point  $z$ . The IMR is sometimes referred to as a ‘hazard ratio.’ This is because a hazard function answers the question ‘what is the probability of an event given that the event has not already occurred.’ Similarly, the IMR answers the question: what is the expectation of epsilon given that epsilon is greater than or equal to  $z$ ?

- We can calculate the expected wage in the source country of workers who do migrate as:

$$\begin{aligned}
E(w_1|\text{Immigrate}) &= \mu_1 + E\left(\varepsilon_1 \middle| \frac{\nu}{\sigma_\nu} > z\right) \\
&= \mu_1 + \rho_{1\nu}\sigma_1 \left(\frac{\phi(z)}{\Phi(-z)}\right)
\end{aligned} \tag{4}$$

## 2.2 THREE CASES

- It's convenient to rearrange (3) and (4) to get

$$\begin{aligned} E(w_0|\text{Immigrate}) &= \mu_0 + \rho_{0\nu}\sigma_0 \left( \frac{\phi(z)}{1 - \Phi(z)} \right) \\ &= \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu} \left( \rho - \frac{\sigma_0}{\sigma_1} \right) \left( \frac{\phi(z)}{1 - \Phi(z)} \right), \end{aligned} \quad (5)$$

and

$$\begin{aligned} E(w_1|\text{Immigrate}) &= \mu_1 + \rho_{1\nu}\sigma_1 \left( \frac{\phi(z)}{1 - \Phi(z)} \right) \\ &= \mu_1 + \frac{\sigma_0\sigma_1}{\sigma_\nu} \left( \frac{\sigma_1}{\sigma_0} - \rho \right) \left( \frac{\phi(z)}{1 - \Phi(z)} \right), \end{aligned} \quad (6)$$

where  $\rho = \sigma_{01}/\sigma_0\sigma_1$  as above.

- [Algebra check

$$\begin{aligned} \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu} \left( \rho - \frac{\sigma_0}{\sigma_1} \right) \left( \frac{\phi(z)}{1 - \Phi(z)} \right) &= \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu} \left( \frac{\sigma_{01}}{\sigma_0\sigma_1} - \frac{\sigma_0}{\sigma_1} \right) \left( \frac{\phi(z)}{1 - \Phi(z)} \right) \\ &= \mu_0 + \left( \frac{\sigma_{01} - \sigma_0^2}{\sigma_\nu} \right) \left( \frac{\phi(z)}{1 - \Phi(z)} \right). \end{aligned}$$

Notice that

$$\text{cov}(\nu, \varepsilon_0) = \sigma_{0\nu} = E[(\varepsilon_1 - \varepsilon_0)\varepsilon_0] = \sigma_{01} - \sigma_0^2,$$

which we can substitute in:

$$\begin{aligned} &= \mu_0 + \left( \frac{\sigma_{0\nu}}{\sigma_\nu} \right) \left( \frac{\phi(z)}{1 - \Phi(z)} \right) \\ &= \mu_0 + \left( \frac{\sigma_{0\nu}}{\sigma_\nu\sigma_0} \right) \sigma_0 \left( \frac{\phi(z)}{1 - \Phi(z)} \right) \\ &= \mu_0 + \rho_{0\nu}\sigma_0 \left( \frac{\phi(z)}{1 - \Phi(z)} \right). \end{aligned}$$

So, this substitution works.]

- Define  $Q_0 = E(\varepsilon_0|I = 1)$ ,  $Q_1 = E(\varepsilon_1|I = 1)$ . We now have three cases:

### 2.2.1 POSITIVE HIERARCHICAL SORTING

- This is a case where immigrants are positively selected from the source country distribution and are also above the mean of the host country distribution:  $Q_0 > 0, Q_1 > 0$ . This will be true iff

$$\frac{\sigma_1}{\sigma_0} > 1 \text{ and } \rho > \frac{\sigma_0}{\sigma_1}.$$

- What do these conditions mean? First,  $\frac{\sigma_1}{\sigma_0} > 1$  implies that the host country has a higher ‘return to skill’ than the source country. Second,  $\rho > \frac{\sigma_0}{\sigma_1}$ , implies that the correlation between the skills valued in the host and source country is sufficiently high. If you were a skilled worker in the source country, you would not want to migrate to a host country with a very high return to skills if the skills valued in the host country were uncorrelated (or negatively correlated) with skills value in the home country.
- 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.

### 2.2.2 NEGATIVE HIERARCHICAL SORTING

- This is a case where immigrants are negative selected from the source country distribution and are also below the average of the host country distribution:  $Q_0 < 0, Q_1 < 0$ . This will be true iff

$$\frac{\sigma_0}{\sigma_1} > 1 \text{ and } \rho > \frac{\sigma_1}{\sigma_0}.$$

- This is simply the converse case. Here, the source country is unattractive to low earnings workers because of high wage dispersion. Again assuming that wages are sufficiently correlated between the source and host country, low skill workers will want to migrate



to take advantage of the ‘insurance’ provided by a narrow wage structure in the host country.

- So, this is the potentially unattractive case (certainly from Borjas’ perspective) where a compressed wage structure ‘subsidizes’ low skill workers, thus attracting low skill workers from abroad.

### 2.2.3 ‘REFUGEE’ SORTING

- A third case is where  $Q_0 < 0, Q_1 > 0$ , that is, immigrants are selected from the lower tail of the home country distribution but arrive in the upper tail of the host country distribution. This can only occur if

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

meaning that the correlation between earnings in the two countries is sufficiently low (could be negative).

- 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).

### 2.2.4 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.
- Mathematically, this case would require that

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

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

### 3 RELEVANCE?

- How relevant is the Borjas/Roy selection model to the problem he studies in the 1987 paper, self-selection of immigrants? The evidence is not overwhelming. There are probably more relevant applications of this model.
  1. One is Borjas' recent NBER paper on self-selection into government employment in the U.S. Though the returns to skills/education rose rapidly in the U.S. during the 1980s, the wage structure in government jobs was stable, meaning that in relative terms, the public sector wage structure became compressed. You would therefore expect a negative hierarchical sorting case to become potentially relevant: high skill workers leave the government for the private sector and low skill workers remain in government jobs to be sheltered from falling wages. This is what Borjas' paper purports to show, and casual empiricism suggests that this is likely to be right.
  2. A closely related application, but one I've not applied, is the changing selection of workers into the teaching profession. The teaching profession probably used to offer relatively high 'returns to skills' for educated women. For many women, it was the only profession open. This has changed considerably in the U.S. Teaching now probably provides a high floor and a low ceiling for wages for college educated females. This would suggest growing adverse selection into teaching over time. Worth exploring.
- The enduring contribution of Borjas' paper for labor economists is its simple and useful formulation of the Roy model. (Note that one limitation of this model is that it ignores general equilibrium effects whereby large immigrant flows would actually change the wage structure parameters in the source and host countries.)
- The growing focus of empirical economists on applying instrumental variables to causal estimation is in large part a response to the realization that self-selection (i.e., optimizing behavior) plagues interpretation of ecological relationships. Hence, understanding the importance of self-selection has vastly improved empirical work.

- But instrumental variables are not the only answer to testing cause and effect with observed data. Self-selection also points to the existence of *equilibrium relationships* that should be observed in ecological data (i.e., those who immigrate should in general do better in the host country than the source country), and these can be tested without an instrument. In fact, there are some natural sciences that proceed almost entirely without experimentation – for example, astrophysics. How do they do it? Models predict non-obvious relationships in data. These implications can be verified or refuted by data, and this evidence strengthens or overturns the hypotheses. Many economists seem to have forgotten this methodology.