Firms and workers II

Stefano Caria

Michaelmas Term 2026

Roadmap

Selection

The Roy model

Lagakos and Waugh (2013): selection into sectors

References

Occupational choice with heterogeneity in more than one dimension

What if people differ not only in entrepreneurial ability but also in skill as workers?

Two-dimensional heterogeneity raises issues of comparative advantage.

Sorting into workers and entrepreneurs is now more complicated.

Fortunately, a useful framework comes from a model of selection that can be traced to Roy (1951).

Roadmap

Selection

The Roy model

Lagakos and Waugh (2013): selection into sectors

References

Roadmap

Selection

The Roy model

Lagakos and Waugh (2013): selection into sectors

References

A model of selection based on Roy (1951)

Focuses on sorting and self-selection into occupational categories. However, the same model structure is used in many contexts and in many fields of economics.

- Main idea is that people sort on the basis of comparative advantage.
- Let's first set up the model in its simplest form — following the presentation in Borjas 1989 — and then look at an application.

The Roy model: setup

Two sectors, occupations, or migration outcomes: 0 and 1.

Each person i has a potential (log) earning in each sector:

$$\log w_{0i} = \mu_0 + \varepsilon_{0i}$$
$$\log w_{1i} = \mu_1 + \varepsilon_{1i}$$

- μ_j: mean log earnings that the average person would earn in sector j
- ε_{ji}: person-specific deviation, i.e. i's skill in sector j
- $(\varepsilon_0, \varepsilon_1)$ jointly normal, mean zero, variances σ_0^2, σ_1^2, correlation ρ
  - σ_0, σ_1: how much each sector *rewards* skill.
  - ρ how *transferable* skill is across sectors.
- There is an entry cost C to enter Sector 1.

The Roy model: the key empirical problem

$$\log w_{0i} = \mu_0 + \varepsilon_{0i}$$
$$\log w_{1i} = \mu_1 + \varepsilon_{1i}$$

We only observe w_{ji} in the sector i actually chooses.

This makes it hard to recover $\mu_1 - \mu_0$.

The Roy model: sector choice

Person i chooses sector 1 if it pays more (net of entry costs). Define the index

$$I_i \;=\; \log\left(\frac{w_{1i}}{w_{0i}+C}\right) \;\approx\; (\mu_1-\mu_0-\pi)+(\varepsilon_{1i}-\varepsilon_{0i}), \qquad \pi \equiv C/w_{0i}$$

derivation

(Note assumption that π is the same for everyone)

- Choose sector 1 iff $I_i > 0$, i.e. iff

$$\varepsilon_{1i}-\varepsilon_{0i} \;>\; \mu_0-\mu_1+\pi \tag{1}$$

- Define $\nu_i \;\equiv\; \varepsilon_{1i}-\varepsilon_{0i}$, and $k \;\equiv\; \mu_0-\mu_1+\pi$. So $I_i > 0$ iff $\nu_i > k$.
- ν is normal with mean zero and variance $\sigma_\nu^2 = \sigma_0^2+\sigma_1^2-2\rho\sigma_0\sigma_1$.

How many people will choose sector 1?

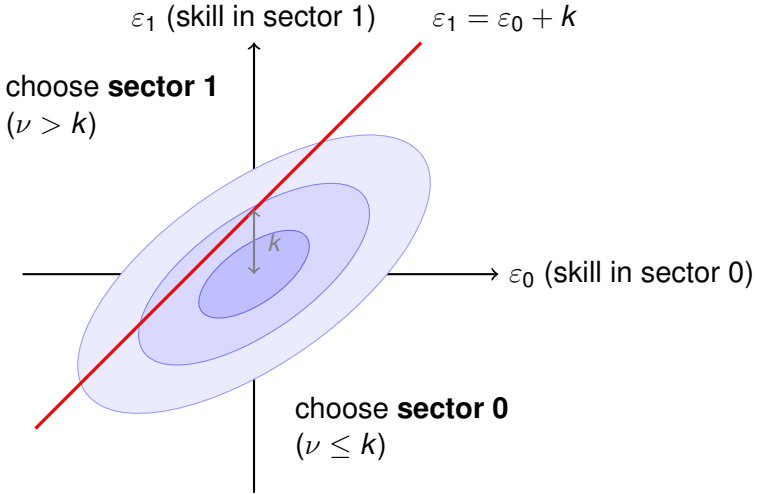
- Share of people in sector 1:

$$\begin{aligned} P = \Pr(\nu > k) &= 1 - \Phi(k/\sigma_\nu) \\ &= 1 - \Phi(z) \end{aligned}$$

- where $z \equiv k/\sigma_\nu$.

- P rises with $\mu_1 - \mu_0$ and falls with π.

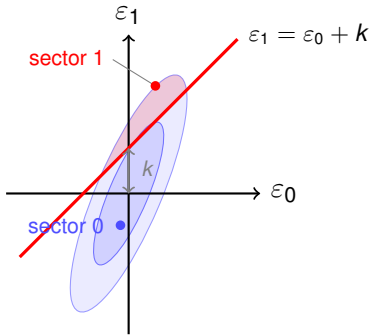
Who selects into sector 1? Graphical analysis



- Everyone above the line goes to sector 1.

Positive selection: who ends up in each sector?

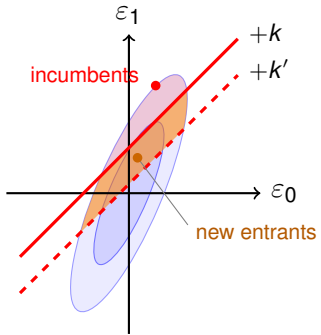
$\sigma_1 > \sigma_0$, $\rho > \sigma_0/\sigma_1$: transferable skill, sector 1 rewards it more



- Sector 1 gets people better at sector 1 *and* better at sector 0.
- Sector 1 attracts them because it rewards high skills strongly ($\sigma_1 > \sigma_0$).

$\rightarrow$ The observed earnings gap *overstates* $\mu_1 - \mu_0$.

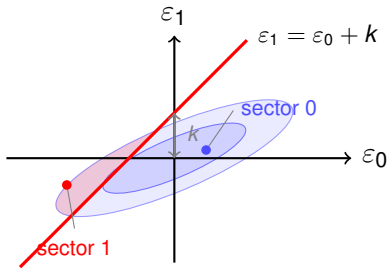
Positive selection: what if we subsidise entry?



- The marginals are *worse* than the incumbents.
- Average quality in sector 1 *falls* as the sector expands.

Negative selection: who ends up in each sector?

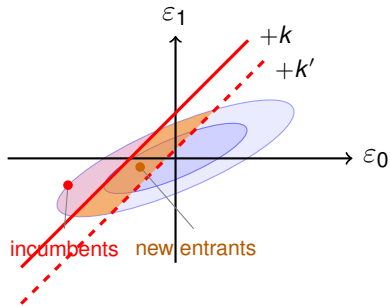
$\sigma_0 > \sigma_1, \quad \rho > \sigma_1/\sigma_0$: transferable skill, sector 0 rewards it more



- Sector 1 gets people worse at sector 1 and worse at sector 0.
- Sector 1 attracts them because it penalises low skills weakly ($\sigma_1 < \sigma_0$).

$\rightarrow$ The observed earnings gap *understates* $\mu_1 - \mu_0$.

Negative selection: what if we subsidise entry?



- Now the marginals are *better* than the incumbents.
- Average quality in sector 1 *rises* as the sector expands.

Aside: conditional means of a truncated bivariate normal

Let x and y be jointly normal with mean zero, and let k be a scalar. Then

$$E[x \mid y > k] = \frac{\text{cov}(x, y)}{\sigma_y} \lambda(z), \qquad z \equiv k/\sigma_y, \quad \lambda(z) \equiv \frac{\phi(z)}{1 - \Phi(z)} \tag{2}$$

In the case of our model, we will be interested in $E[\varepsilon_j \mid I > 0] = E[\varepsilon_j \mid \nu > k]$. Recall $\nu_i \equiv \varepsilon_{1i} - \varepsilon_{0i}$.

We can apply the result in equation (2) to find $E[\varepsilon_j \mid \nu > k]$. Note that :

$$\text{cov}(\varepsilon_0, \nu) = \text{cov}(\varepsilon_0, \varepsilon_1 - \varepsilon_0) = \rho\sigma_0\sigma_1 - \sigma_0^2, \qquad \text{cov}(\varepsilon_1, \nu) = \sigma_1^2 - \rho\sigma_0\sigma_1$$

Who selects into sector 1? The conditional means

The average skill of those who choose sector 1, in *each* sector:

$$Q_0 \equiv E[\varepsilon_0 \mid I > 0] = \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\rho - \frac{\sigma_0}{\sigma_1}\right)\lambda(z)$$

$$Q_1 \equiv E[\varepsilon_1 \mid I > 0] = \frac{\sigma_0\sigma_1}{\sigma_\nu}\left(\frac{\sigma_1}{\sigma_0} - \rho\right)\lambda(z)$$

where $\lambda(z) = \phi(z)/[1 - \Phi(z)] > 0$ is the inverse Mills ratio.

- Q_1 tells us whether sector 1 attracts people who are good *at sector 1*;
- Q_0 tells whether it attracts people who would have been good *at sector 0*;
- The sign of both only depends on ρ (the transferability of skills) and $\frac{\sigma_1}{\sigma_0}$ (which sector rewards skills more).

Three types of selection (Borjas, 1987)

	Q_0	Q_1	Condition
Positive selection	> 0	> 0	$\sigma_1 > \sigma_0$ and $\rho > \sigma_0/\sigma_1$
Negative selection	< 0	< 0	$\sigma_0 > \sigma_1$ and $\rho > \sigma_1/\sigma_0$
"Refugee" sorting	< 0	> 0	$\rho < \min\{\sigma_0/\sigma_1, \sigma_1/\sigma_0\}$
(Impossible)	> 0	< 0	would require $\rho > \max\{\sigma_0/\sigma_1, \sigma_1/\sigma_0\} \geq 1$

Implication (1): observed gaps are not causal effects

Compare average log earnings of those *observed* in each sector:

$$E[\log w_1 \mid I > 0] - E[\log w_0 \mid I \leq 0] = \underbrace{(\mu_1 - \mu_0)}_{\text{“treatment” effect}} + \underbrace{E[\varepsilon_1 \mid I > 0] - E[\varepsilon_0 \mid I \leq 0]}_{\text{selection}}$$

- Even if $\mu_1 = \mu_0$ (no return to moving), the observed gap can be of either sign.
- This issue arises in many development applications:
  - urban–rural wage gaps,
  - entrepreneur–worker earnings gaps,
  - formal–informal gaps,
  - sectoral productivity gaps (next).
- This is precisely the mechanism explored in Lagakos and Waugh (2013), which we will turn to next.

Implication (2): composition changes when the margin moves

Suppose we subsidise entry into sector 1 ($\pi \downarrow$) or that sector 1 becomes more attractive ($\mu_1 \uparrow$) :

- z falls and more people enter.
- As $z \downarrow$, $\lambda(z) \downarrow$:
  - Under positive selection, average quality in sector 1 *falls* as it expands ($Q1$ falls)

Roadmap

Selection

The Roy model

Lagakos and Waugh (2013): selection into sectors

References

Cross-country labor productivity differences are much larger in agriculture compared to non-agriculture

Labour productivity measured as output per worker

TABLE 1—SECTOR LABOR PRODUCTIVITY DIFFERENCES AND EMPLOYMENT SHARES

	Agriculture	Aggregate	Non-agriculture	Ag/non-agriculture ratio
90–10 labor productivity difference	45	22	4	10.7
Employment shares ninetieth percentile country	3		97	
Employment shares tenth percentile country	78		22	

Notes: The aggregate productivity difference is the ratio of GDP per worker between the ninetieth and tenth percentile countries. Sector productivity differences are the ratio of sector output per worker in the ninetieth and tenth percentile countries. The Ag/Non-agriculture Ratio is the agriculture productivity difference divided by the non-agriculture productivity difference.

Source: Caselli (2005).

Do these difference in output per worker (Y/L) reflect genuine differences in technology (A)? Is LMIC's technology gap in agriculture greater than their technology gap in non-agriculture?

Lagakos and Waugh (2013): Understanding this point is crucial!

Developing countries have most of their workers in agriculture.

Thus, their low productivity in agriculture accounts for nearly all of their low productivity in the aggregate.

$\rightarrow$ understanding why productivity differences in agriculture are so large compared to those in the non-agricultural sector is key to explain world income inequality.

Lagakos and Waugh (2013)

They use a Roy model to think about occupational sorting between sectors – agriculture and non-agriculture.

Ingredients:

- Two sectors: agriculture (a) and non-agriculture (n).
- Countries differ in "economy-wide efficiency" A (relative efficiency in agriculture is the same as relative efficiency in non-agriculture)
- Subsistence requirements in preference for ag goods (e.g., Stone-Geary!) that make the demand for agriculture very high when productivity is low.
- Workers heterogeneous in productivity in each sector (Roy, 1951)

Key idea: self-selection of heterogeneous workers determines sector productivity

In poor countries, where economy-wide efficiency is low, most people must work in the agricultural sector in order to satisfy subsistence consumption needs. This is what Schultz (1953) famously called the “food problem.”

Insight: precisely because majority of workers in poor countries employed in agriculture, many must be relatively unproductive at agricultural work.

In rich countries, where A is high, those few workers selecting into agriculture must be those who are relatively most productive at agricultural.

Thus, two countries that differ in economy-wide efficiency will have even larger measured differences in agricultural productivity.

What about productivity in non-agriculture?

Let's now look at the model more formally.

What about productivity in non-agriculture?

Let's now look at the model more formally.

Households

- **Preferences**
  $U^i = \log(c_a^i - \bar{a}) + \nu \log(c_n^i)$

- **Budget constraint**
  $p_a c_a^i + c_n^i \leq y^i$

- Each household has an endowment of productivity/effective units of labor:
  $\{z_a^i, z_n^i\}$

- $\{z_a^i, z_n^i\}$ drawn from distribution $G(z_a, z_n)$

Production

- Production in the economy is characterized by an economy-wide level of productivity A.
- Each sector has a simple aggregate production technology:
  $Y_a = AL_a$ and $Y_n = AL_n$

Because we are in a constant returns to scale world, we do not need to worry about individual firms, and we can just look at the aggregate production.

Production

- From the perspective of the economy as a whole, each sector uses an effective labor input that is just the aggregation of the effective units supplied by individual workers:

$$L_a \equiv \int_{i \in \Omega_a} z_a^i \, dG_i \qquad \text{and} \quad L_n \equiv \int_{i \in \Omega_n} z_n^i \, dG_i$$

- Ω_a (Ω_n) is the set of workers that work in ag (non-ag).

- Note that the aggregate labour units will be different from the number of workers. The number of workers in each sector is given by:

$$N_a \equiv \int_{i \in \Omega_a} dG_i \qquad \text{and} \quad N_n \equiv \int_{i \in \Omega_n} dG_i$$

Sector Choice and Labor Income

- Due to competition: $w_a = p_a A$ and $w_n = A$ (w is wage per effective unit of labor)

- How do workers sort into sectors? They compare labour earnings in each sector. This is simply:

$$y_i \equiv \max\{p_a A z_a^i,\ A z_n^i\}$$

- Workers will work in non-agriculture if and only if:

$$\frac{z_i^n}{z_i^a} \geq p_a$$

Equilibrium

An equilibrium is:

- Relative food price p_a
- Wages w_a and w_n
- Allocations to sectors Ω_a and Ω_n

such that (i) workers optimize, and (ii) labor markets and output markets clear.

This model generates two key propositions.

(1) Relative Price of Food Higher in Poor Countries

Proposition: Consider two economies, rich and poor, with efficiency terms A^R and A^P such that $A^R > A^P$. In equilibrium, the relative price of agriculture is higher in the poor economy: $p_a^P > p_a^R$.

Intuition:

- Poor country demands relatively more food because of subsistence needs
- To induce workers to enter agriculture, given that they are relatively low-skill at agriculture, the only way to persuade them to join the agricultural workforce is for the price of food to be sufficiently high.
- So this economy will "need" a higher p_a in order to reach equilibrium.

(2) Relative Ag Productivity Lower in Poor Countries

Proposition: Consider two economies such that $A^R > A^P$. Assume $E[z_a|z_a/z_n > x]$ and $E[z_n|z_n/z_a > x]$ increase in x. Then equilibrium labour productivity is such that:

$$\frac{Y_a^R/N_a^R}{Y_a^P/N_a^P} > \frac{A^R}{A^P} \tag{3}$$

$$\frac{Y_n^R/N_n^R}{Y_n^P/N_n^P} < \frac{A^R}{A^P} \tag{4}$$

Intuition

$$\frac{Y_a^R/N_a^R}{Y_a^P/N_a^P} > \frac{A^R}{A^P} \tag{5}$$

- In rich country A is higher, and hence p_a is lower (prop 1).
- Lower returns to ag $\rightarrow$ only workers with high comparative advantage select in agriculture in rich country.
- If workers with greater comparative advantage also have higher absolute advantage, then agricultural productivity in rich country increases.

Simulation with economy-wide efficiency difference: selection alone can account for a large share of the differential productivity of ag

TABLE 2—90–10 PRODUCTIVITY DIFFERENCES, DATA AND BENCHMARK MODEL

	Agriculture	Aggregate	Non-agriculture	Ag/non-agriculture ratio
Data	45	22	4	10.7
Model	29	22	13	2.2
Without selection	19	19	19	1.0

Notes: The aggregate productivity difference is the ratio of GDP per worker between the ninetieth and tenth percentile countries. Sector productivity differences are the ratio of sector output per worker in the ninetieth and tenth percentile countries. The ag/non-agriculture ratios are the agriculture productivity differences divided by the non-agriculture productivity differences.

Source: Authors' calculations and Caselli (2005).

Roadmap

Selection

The Roy model

Lagakos and Waugh (2013): selection into sectors

References

Key references

- (*) Borjas, George J. ”Self-selection and the earnings of immigrants” The American Economic Review (1987).
  Section I (i.e., focus on the theory, not the empirical part of the paper)

- (*) Lagakos and Waugh (2013) Selection, agriculture, and cross-country productivity differences. American Economic Review 103, no. 2 (2013): 948-980.
  Theory is key, but have a quick look at the empirics too.

Derivation back

Note that:

$$\begin{aligned}\log\left(\frac{w_{1i}}{w_{0i}+C}\right) &= \log w_{1i} - \log\left(w_{0i} + C\right) \\ &= \log w_{1i} - \log\left(w_{0i}\left(1 + C/w_{0i}\right)\right) \\ &= \log w_{1i} - \log w_{0i} - \log\left(1 + \frac{C}{w_{0i}}\right)\end{aligned}$$

Then, note that, for x small, $\log(1+x) \approx x$. So $\log(1 + C/w_{0i}) \approx C/w_{0i}$.
Define $\pi_i \equiv C/w_{0i}$ and substitute $\log w_{ji} = \mu_j + \varepsilon_{ji}$. This gives:

$$l_i \approx (\mu_1 + \varepsilon_{1i}) - (\mu_0 + \varepsilon_{0i}) - \pi_i = (\mu_1 - \mu_0 - \pi_i) + (\varepsilon_{1i} - \varepsilon_{0i})$$