Assignment: Programming and Roy Model

Due: September 20, 2022

1 Programming Setup

1.1 Setup DataCamp

1. Log into the course platform
2. Finish an arbitrary course

1.2 R

1. Install R
2. Install RStudio
3. Install Packages: `data.table, ggplot2, fixest, knitr, dplyr`
4. Generate RMarkdown files and knit it: "HelloWorld.rmd" and "HelloWorld.pdf"

1.3 Debugger

1. Write a function that adds up two numbers
2. Initiate a debugger and examine what the value

1.4 Setup Github

1. Sign up an Github account
2. Create an (open) repository, named "ECON-XXXX"
3. Setup RStudio to connect with Github

4. Push the "HelloWorld.rmd" and "HelloWorld.pdf" you generated

2 Sign up NBER working paper series

1. What is the title of the second paper listed on the NBER weekly working paper series (either from Sep. 5, Sep. 12, or Sep. 19)

2. Identify and download a paper that interests you

3 Sign up SRDA

1. Register for an SRDA account

2. Download an arbitrary year of PSFD

3. Which year did you download?

4. Plot rate of working against age. (x-axis: mean of work dummy, y-axis: age) using `ggplot2`

4 Roy Model

4.1 Review

Consider the simple model:

$$w_0 = \mu_0 + \epsilon_0$$

$$w_1 = \mu_1 + \epsilon_1$$

where migrant is 1, non-migrant is 0. Assume that $\epsilon_0 \sim N(0, \sigma_0^2)$ and $\epsilon_1 \sim N(0, \sigma_1^2)$. Assume the migrant cost C, and the correlation coefficient is $\rho = \frac{\sigma_{01}}{\sigma_0 \sigma_1}$. One will migrate if $w_1 > w_0 + C$. Let $\nu = \epsilon_1 - \epsilon_0$. Let $z = \frac{\mu_0 - \mu_1 + C}{\sigma_\nu}$.

1. Show that

$$E[w_0|I] = \mu_0 + \frac{\sigma_0 \sigma_1}{\sigma_\nu}(\rho - \frac{\sigma_0}{\sigma_1})(\frac{\phi(z)}{1 - \Phi(z)}) \tag{1}$$

$$E[w_1|I] = \mu_1 + \frac{\sigma_0 \sigma_1}{\sigma_\nu}(\frac{\sigma_1}{\sigma_0} - \rho)(\frac{\phi(z)}{1 - \Phi(z)}) \tag{2}$$

2. Let $Q_0 = E[\epsilon_0|I]$ and $Q_1 = E[\epsilon_1|I]$. Is it possible that $Q_0 > 0, Q_1 < 0$?

4.2 Simulation

1. Pick your favorite value for this set of parameters $(\mu_0, \mu_1, \sigma_0, \sigma_1, \sigma_{01}, C)$
2. Simulate the (ϵ_0, ϵ_1) for N equals to 10 million individuals. Use a `data.table` to store these.
3. Create the columns for w_0 and w_1
4. Generate the column I that take binary value.
5. Calculate $E[w_0|I]$, $E[w_1|I]$, Q_0, Q_1 from data without invoking equation (1) or (2).
6. Calculate RHS of equation (1) and (2) to compare with the previous question.
7. Which columns are observed in the real world? Which of them are not?

5 Roy Model is Everywhere

1. Find an example in applied economics. (Your master thesis, your friend's thesis, your previous term paper, etc.)
2. Use the Roy model notation to write down a (simple version) of the research question
3. Explain this to (any) other people use both the intuition and the mathematical framework. (I'll randomly draw two people to teach us on Sep. 20th's class.)