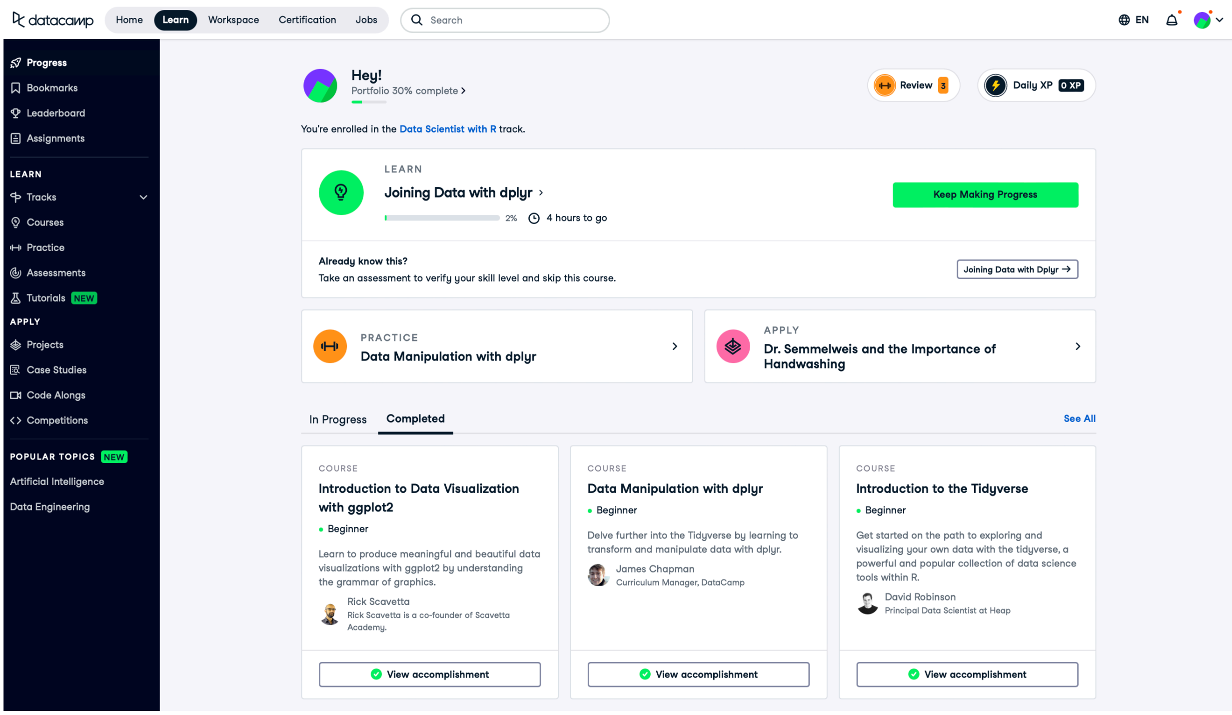
National Taiwan University
ECON 5211 Labor Economics I
Prof. Kuan Ming, Chen

Xiang Jyun, Jhang
Homework 1
March 10, 2024

1 Programming Setup

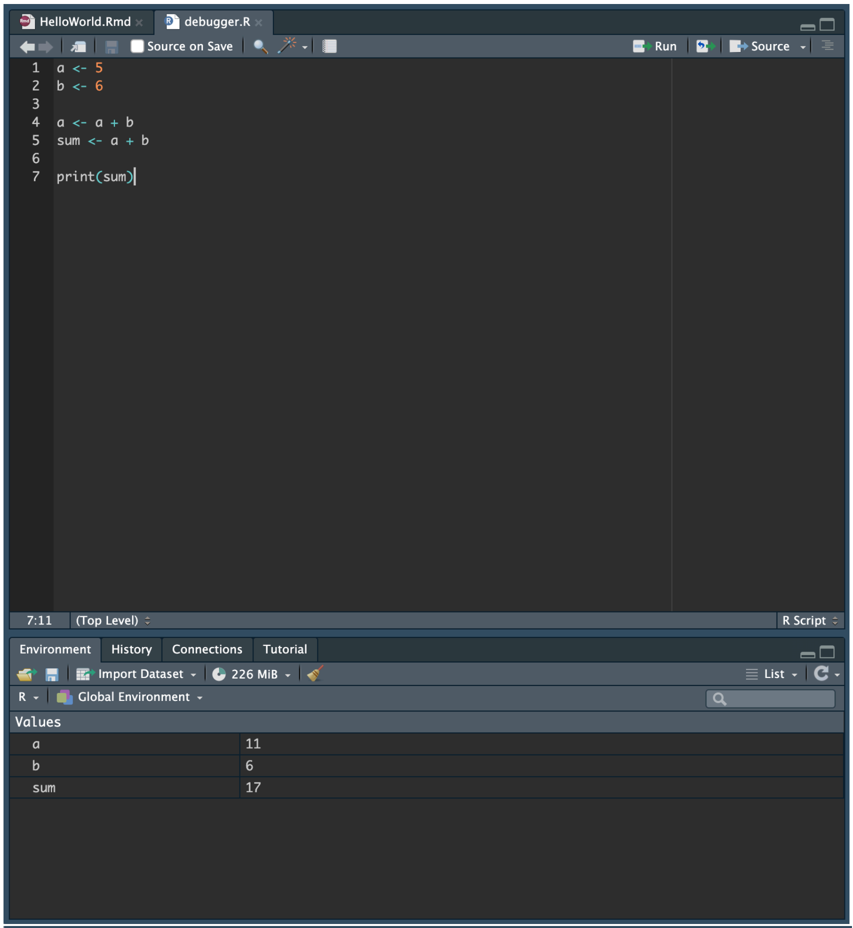
1.1 Setup Datacamp

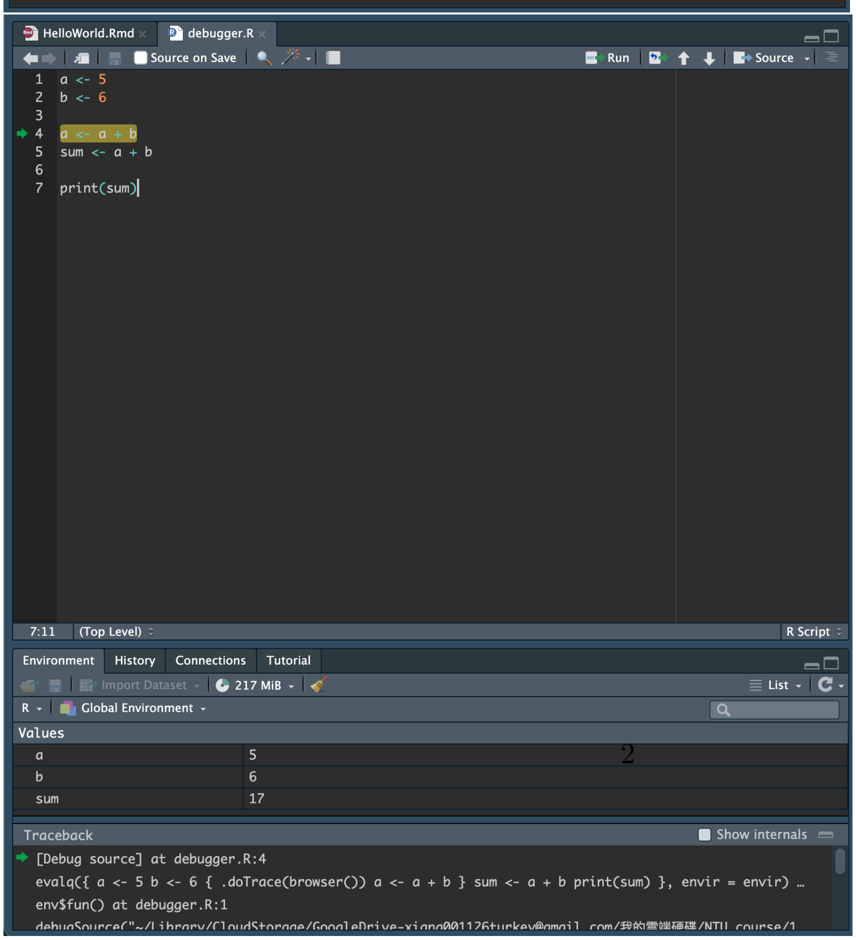


1.2 R

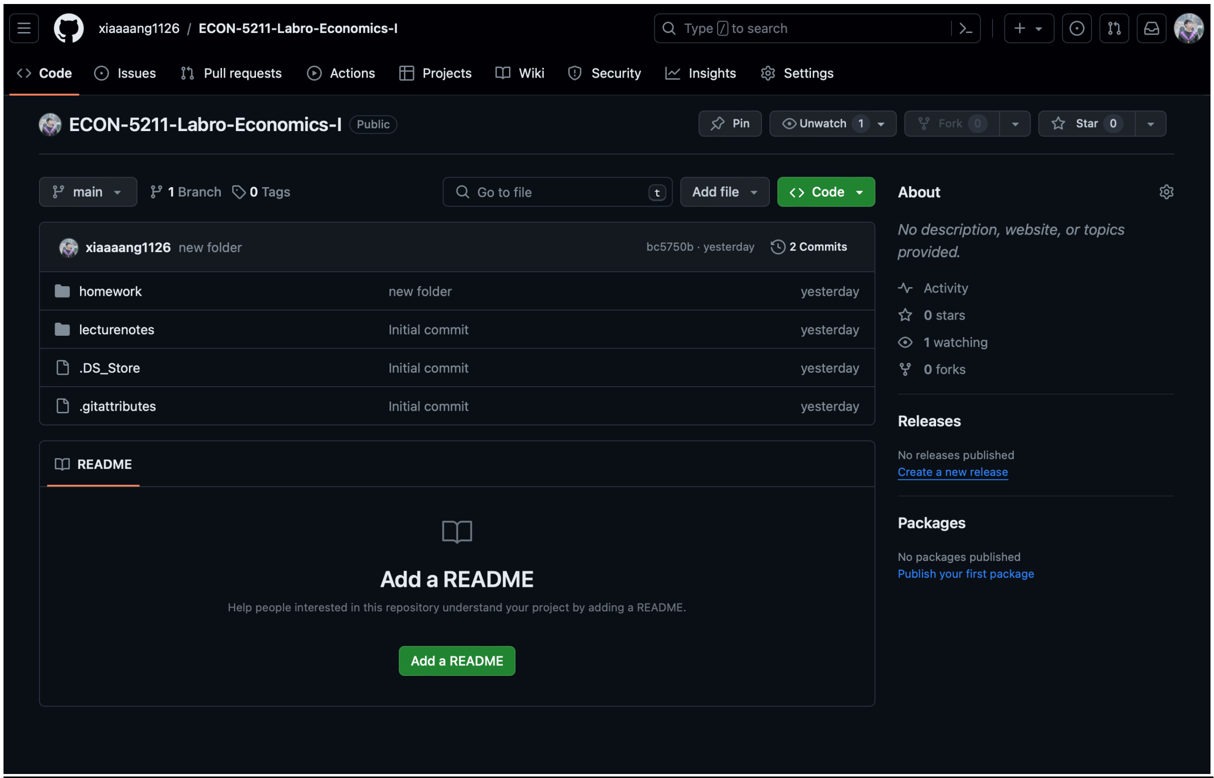
For installation process and HelloWorld example, please refer to the file located at `"/homework/homework1/R_code"`

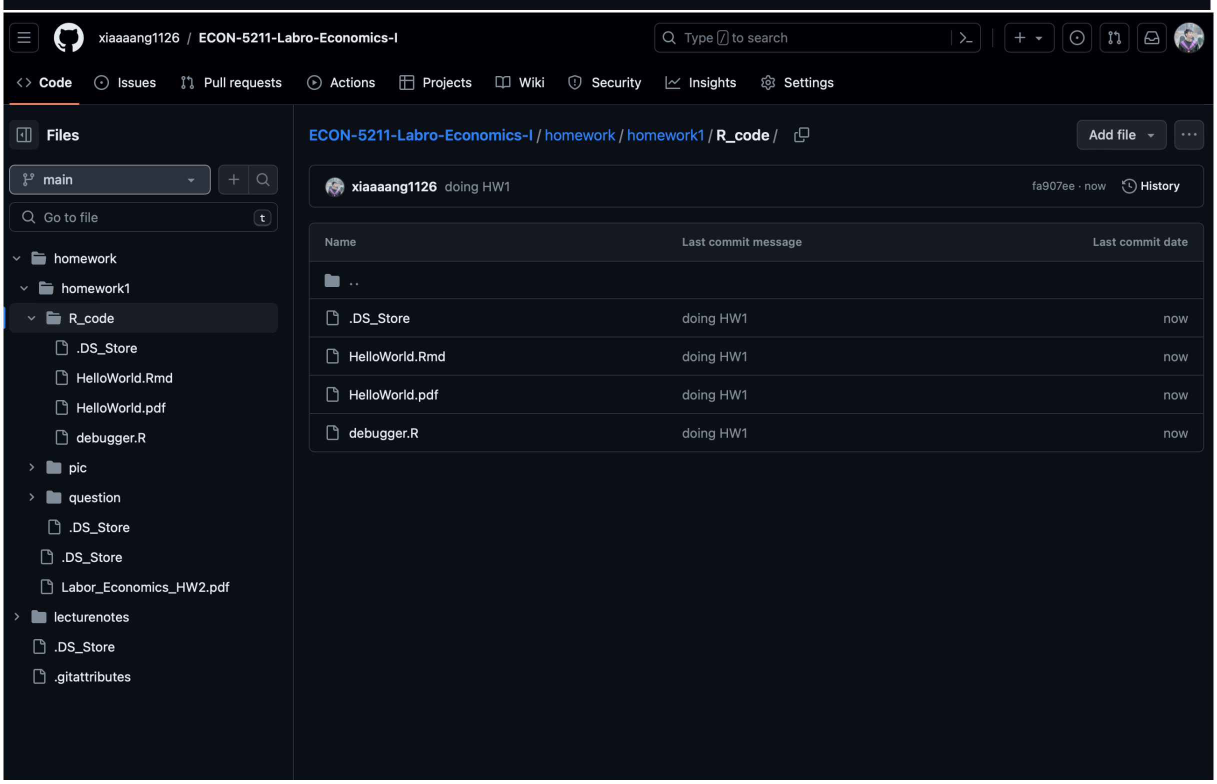
1.3 Debugger


HelloWorld.Rmd
debugger.R
Source on Save
Run
Source
1 a <- 5
2 b <- 6
3
4 a <- a + b
5 sum <- a + b
6
7 print(sum)
7:11
(Top Level)
R Script
Environment
History
Connections
Tutorial
Import Dataset
226 MiB
List
R
Global Environment
Values
a 11
b 6
sum 17


HelloWorld.Rmd
debugger.R
Source on Save
Run
Source
1 a <- 5
2 b <- 6
3
4 a <- a + b
5 sum <- a + b
6
7 print(sum)
7:11
(Top Level)
R Script
Environment
History
Connections
Tutorial
Import Dataset
217 MiB
List
R
Global Environment
Values
a 5
b 6
sum 17
Traceback
Show internals
[Debug source] at debugger.R:4
evalq({ a <- 5 b <- 6 { .doTrace(browser()) a <- a + b } sum <- a + b print(sum) }, envir = envir) …
env$fun() at debugger.R:1

1.4 Setup Github





2 Sign up NBER working paper series

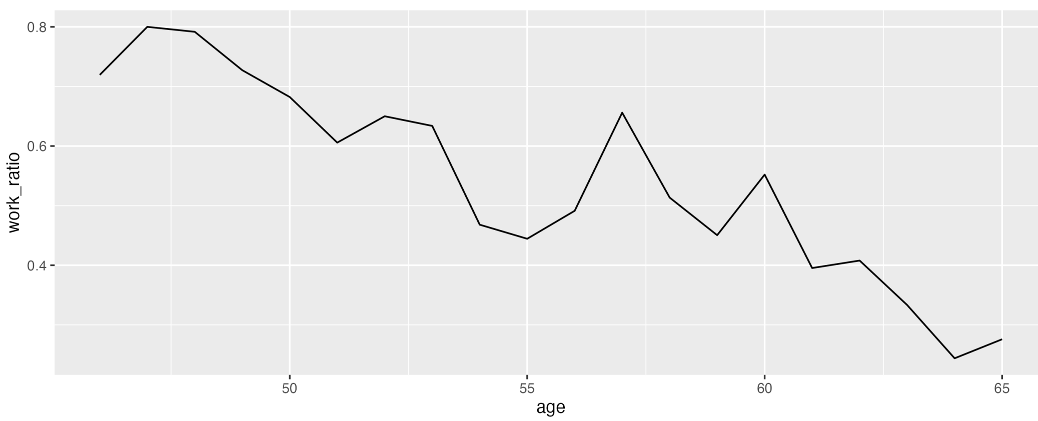
1. The title of the second paper listed on the NBER weekly working paper series is *"Why Survey-Based Subjective Expectations are Meaningful and Important"*, working paper number 32199

2. I have downloaded a paper interesting me most at the path `"/homework/homework1/path"` with title *"Why Survey-Based Subjective Expectations are Meaningful and Important"*

3 Sign up SRDA

1. The year of the PSFD dataset I downloaded is 2000

2. The working rate against age in the dataset on 2000 year is



4 Roy Model

4.1 Review

1. We first consider $\mathbb{E}[w_0|\text{Migrate}]$, which can be expressed as

$$\begin{aligned}\mathbb{E}[w_0|\text{Migrate}] &= \mu_0 + \mathbb{E}[\epsilon_0|w_1 > w_0 + C] \\ &= \mu_0 + \mathbb{E}[\epsilon_0|\mu_1 + \epsilon_1 > \mu_0 + \epsilon_0 + C] \\ &= \mu_0 + \mathbb{E}[\epsilon_0|\nu > \mu_0 - \mu_1 + C] \\ &= \mu_0 + \mathbb{E}[\epsilon_0|\frac{\nu}{\sigma_\nu} > z] \\ &= \mu_0 + \sigma_0\mathbb{E}[\frac{\epsilon_0}{\sigma_0}|\frac{\nu}{\sigma_\nu} > z]\end{aligned}$$

next, to compute the conditional expectation of ϵ_0, we have to derive the following formulas

(a) Under normality, we have the regression coefficient formula

$$\mathbb{E}[\epsilon_0|\nu] = \frac{\sigma_{0,\nu}}{\sigma_\nu^2}\nu$$

(b) Use the formula after changing terms

$$\mathbb{E}[\frac{\epsilon_0}{\sigma_0}|\frac{\nu}{\sigma_\nu}] = \frac{1}{\sigma_0}\underbrace{\frac{\sigma_{0,\nu}}{\sigma_\nu^2}\frac{\nu}{\sigma_\nu}}_{\substack{\text{by above}\\ \text{formula}}}\frac{1}{\sigma_\nu^{-2}}\frac{1}{\sigma_\nu} = \frac{\sigma_{0,\nu}}{\sigma_0\sigma_\nu}\frac{\nu}{\sigma_\nu} = \rho_{0,\nu}\frac{\nu}{\sigma_\nu} \tag{1}$$

(c) From the preceding expectation formula, we can simplify the conditional expectation

$$\begin{aligned}\mathbb{E}[w_0|\text{Migrate}] &= \mu_0 + \sigma_0\mathbb{E}[\frac{\epsilon_0}{\sigma_0}|\frac{\nu}{\sigma_\nu} > z]\\ &\overset{(1)}{=} \mu_0 + \rho_{0,\nu}\sigma_0\mathbb{E}(\frac{\nu}{\sigma_\nu}|\frac{\nu}{\sigma_\nu} > z)\\ &= \mu_0 + \rho_{0,\nu}\sigma_0\frac{\phi(z)}{1-\Phi(z)}\end{aligned}$$

where $\Phi(z)$ and $\phi(z)$ are respectively defined as the pdf and cdf of z.

Besides, note that the coefficient of Inverse Mill Ratio can be rewritten as

$$\rho_{0,\nu}\sigma_0 = \underbrace{\frac{\sigma_{0,\nu}}{\sigma_\nu} = \frac{\sigma_{0,1}-\sigma_0^2}{\sigma_\nu}}_{\substack{\sigma_{0,\nu}=\text{Cov}(\epsilon_0,\epsilon_1-\epsilon_0)\\ =\sigma_{0,1}-\sigma_0^2}} = \frac{\sigma_0\sigma_1}{\sigma_\nu}(\frac{\sigma_{0,1}-\sigma_0^2}{\sigma_0\sigma_1}) = \frac{\sigma_0\sigma_1}{\sigma_\nu}(\rho - \frac{\sigma_0}{\sigma_1})$$

Follow the same procedure, we can derive $\mathbb{E}[w_1|\text{Migrate}]$ as well. Therefore, the conditional expectation can be expressed in a quite symmetric way

$$\begin{cases}\mathbb{E}[w_0|\text{Migrate}] = \mu_0 + \frac{\sigma_0\sigma_1}{\sigma_\nu}(\rho - \frac{\sigma_0}{\sigma_1})(\frac{\phi(z)}{1-\Phi(z)})\\ \mathbb{E}[w_1|\text{Migrate}] = \mu_1 + \frac{\sigma_0\sigma_1}{\sigma_\nu}(\frac{\sigma_1}{\sigma_0} - \rho)(\frac{\phi(z)}{1-\Phi(z)})\end{cases}$$

2. If $Q_0 > 0$ and $Q_1 < 0$, then we have

$$\begin{cases}\rho - \frac{\sigma_0}{\sigma_1} > 0\\ \frac{\sigma_1}{\sigma_0} - \rho < 0\end{cases} \quad\Leftrightarrow\quad \rho = \frac{\sigma_{0,1}}{\sigma_0\sigma_1} > \max\{\frac{\sigma_0}{\sigma_1}, \frac{\sigma_1}{\sigma_0}\} \quad\Leftrightarrow\quad \sigma_{0,1} > \max\{\sigma_0^2, \sigma_1^2\}$$

which is impossible, hence such situation doesn't occur given people choose to migrate. In economics sense, this contradiction suggest that people would not leave for the lower proportion of w_1 income distribution when they are in the upper proportion of w_0 income distribution

4.2 Simulation

1. First, let's pick some values for the parameters $\{\mu_0, \mu_1, \sigma_0, \sigma_1, \sigma_{0,1}, C\}$

```r
# 1. Pick my favorite value for set of parameters
mu_0 <- 100
mu_1 <- 100
sigma_0 <- 3
sigma_1 <- 6
sigma_01 <- 5
cost <- 50
n <- 1000000

```

2. After simulation, store the result of $\{\epsilon_0, \epsilon_1\}$ in `data.table`

```r
# 2. Simulate the epsilon_0 and epsilon_1
set.seed(123)
reps <- 1000000
par.est.DT <- data.table(matrix(0, reps, 6))
epsilon_0 <- rnorm(n, mean = 0, sd = sigma_0**2)
epsilon_1 <- rnorm(n, mean = 0, sd = sigma_1**2)

```

3. and create the columns for w_0 and w_1

```r
# 3. Create the columns for w0 and w1
w_0 <- mu_0 + epsilon_0
w_1 <- mu_1 + epsilon_1
par.est.DT[, 1] <- epsilon_0
par.est.DT[, 2] <- epsilon_1
par.est.DT[, 3] <- w_0
par.est.DT[, 4] <- w_1

```

4. also, generate a column I that takes binary value

```r
# 4. Generate the column I that take binary value.
par.est.DT[, 5] <- ifelse(par.est.DT[, 2] > par.est.DT[, 1], 1, 0)

```

5. Then compute $\mathbb{E}[w_0|\mathbb{I}], \mathbb{E}[w_1|\mathbb{I}], Q_0, Q_1$ from data

```r
# 5. Calculate E[w0|I], E[w1|I], Q0, Q1 from data
E_epsilon0_D1 <- par.est.DT[V5 == 1, mean(V1)]
E_epsilon1_D1 <- par.est.DT[V5 == 1, mean(V2)]
E_w0_D1 <- par.est.DT[V5 == 1, mean(V3)]
E_w1_D1 <- par.est.DT[V5 == 1, mean(V4)]

```

6. Use equation (1) and (2) to compute conditional expectation of w_0 and w_1

```r
# 6. Calculate RHS of equation (1) and (2)
nu <- epsilon_1 - epsilon_0
par.est.DT[, 6] <- nu
sigma_nu <- par.est.DT[,sd(V5)]
Z <- (mu_0 - mu_1 + cost)/sigma_nu
mill <- function(x) {
    pnorm(x, lower.tail=FALSE, log.p=TRUE) - dnorm(x, log=TRUE)
}
```

```
RHS_1 <- mu_0 + (sigma_0*sigma_1/sigma_nu) * (sigma_01/(sigma_0*sigma_1) - sigma_0/sigma_1) * mill(Z)
RHS_2 <- mu_1 + (sigma_0*sigma_1/sigma_nu) * (sigma_0/sigma_1 - sigma_01/(sigma_0*sigma_1)) * mill(Z)

# comparison
E_w0_D1  # 98.24212
E_w1_D1  # 127.8435
RHS_1    # 136.8422
RHS_2    # 63.15782

```

7. The column $\mathbb{I}$ and $\mathbb{E}[w_1|\mathbb{I} = 1]$ are observable in the real world, but the other variable $\{\epsilon_0, \epsilon_1, Q_0, Q_1, \phi(Z), \Phi(Z)\}$ and $\mathbb{E}[w_0|\mathbb{I} = 1]$ are unobservable

5 Roy Model is Everywhere

1. An example that immediately comes to my mind is getting master (or ph.D) degree or not. The time spent on studying a degree or the risk of not finishing it are costly, nonetheless, the degree itself is also desirable when seeking a job. It might need argue that $Q_0 < 0$ or $Q_0 > 0$, but I think most people will agree on $Q_1 > 0$.

2. Consider the Roy model

$$\begin{cases} w_0 = \mu_0 + \epsilon_0 \\ w_1 = \mu_1 + \epsilon_1 \end{cases}$$

where studying a higer degree represents 1, while not studying represents 0. It's assumed that $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, where $i \in \{0, 1\}$. Note that the pain or time studying takes are capture by constant cost C and one choose studying higher degree only if $\mathbb{E}[w_1] - C > \mathbb{E}[w_0]$.