ECON613 HW4

Zhilin Tang, zt53

2022-04-20

```
library(ggplot2)
library(gridExtra)
library(dplyr)
library(data.table)
library(tinytex)
library(tidyr)
library(panelr)
library(stats)
library(stargazer)
```

Exercise 1 Preparing the Data

```
# load dataset
dat_raw = read.csv('dat_A4.csv', header=TRUE)
dat = dat_raw[,-1]
```

1.1 Create additional variable for the age of the agent `age`; total work experience measured in years `work_exp`.

```
# age
dat['age'] = 2019 - dat$KEY_BDATE_Y_1997
# work_exp
worked_week = c(paste0('CV_WKSWK_JOB_DLI.0',1:9,'_2019'),
                'CV_WKSWK_JOB_DLI.10_2019',
                'CV_WKSWK_JOB_DLI.11_2019')
dat = dat %>%
  mutate_at(worked_week,funs(ifelse(is.na(.),0,.))) %>%
  mutate(work_exp=rowSums(select(.,CV_WKSWK_JOB_DLI.01_2019:CV_WKSWK_JOB_DLI.11_2019))/52)
```

1.2 Create additional education variables indicating total years of schooling from all variables related to education in our dataset.

```
# individual's years of education
dat['edu'] = case_when(
  dat$YSCH.3113_2019==1 ~ 0,  # None
  dat$YSCH.3113_2019==2 ~ 4,  # GED
  dat$YSCH.3113_2019==3 ~ 12, # High school diploma
  dat$YSCH.3113_2019==4 ~ 15, # AA
  dat$YSCH.3113_2019==5 ~ 16, # BA, BS
```

```r
  dat$YSCH.3113_2019==6 ~ 18, # MA, MS
  dat$YSCH.3113_2019==7 ~ 22, # PhD
  dat$YSCH.3113_2019==8 ~ 18  # DDS, JD, MD
)
```

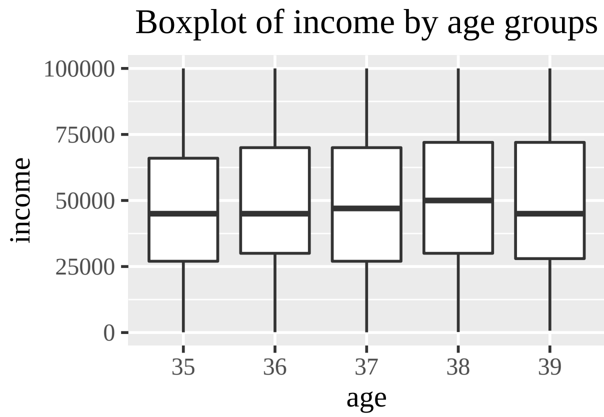
1.3 Provide the following visualizations

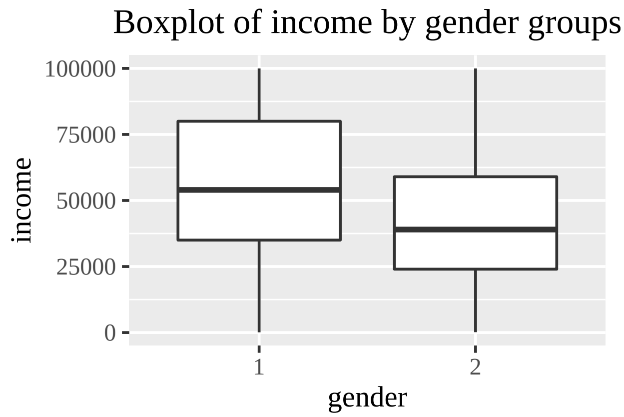
1.3.1 Plot the income data (where income is positive)

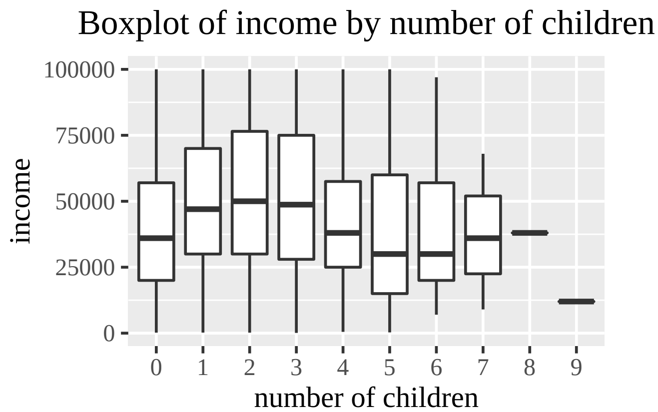
```r
dat131 = dat %>%
  filter(YINC_1700_2019>0)
# i) by age groups
p1 = ggplot(dat131,aes(as.character(age),YINC_1700_2019)) +
  geom_boxplot() +
  xlab('age') +
  ylab('income') +
  ggtitle('Boxplot of income by age groups') +
  theme(plot.title = element_text(hjust = 0.5),text=element_text(family='Times'))

# ii) by gender groups, 1 for male, 2 for female
p2 = ggplot(dat131,aes(as.character(KEY_SEX_1997),YINC_1700_2019)) +
  geom_boxplot() +
  xlab('gender') +
  ylab('income') +
  ggtitle('Boxplot of income by gender groups') +
  theme(plot.title = element_text(hjust = 0.5),text=element_text(family='Times'))

# iii) by number of children
p3 = ggplot(dat131%>%filter(!is.na(CV_BIO_CHILD_HH_U18_2019)),
            aes(as.character(CV_BIO_CHILD_HH_U18_2019),YINC_1700_2019)) +
  geom_boxplot() +
  xlab('number of children') +
  ylab('income') +
  ggtitle('Boxplot of income by number of children') +
  theme(plot.title = element_text(hjust = 0.5),text=element_text(family='Times'))
grid.arrange(p1,p2,p3,ncol=2)
```







1.3.2 Table the share of "0" in the income data

```
dat132 = dat %>%
  filter(YINC_1700_2019>=0)
dat132_0 = dat %>%
  filter(YINC_1700_2019==0)
# i) by age groups
table_income = dat132 %>%
  group_by(age) %>%
  summarise(n = n())
table_income0 = dat132_0 %>%
  group_by(age) %>%
  summarise(n0 = n())
cbind(age=table_income$age,
      share_of_0=table_income0$n0/table_income$n)
```

```
##      age  share_of_0
## [1,]  35 0.009293680
## [2,]  36 0.006300630
## [3,]  37 0.005420054
## [4,]  38 0.008960573
## [5,]  39 0.002994012
```

```r
# ii) by gender groups
table_gender = dat132 %>%
  group_by(KEY_SEX_1997) %>%
  summarise(n = n())
table_gender0 = dat132_0 %>%
  group_by(KEY_SEX_1997) %>%
  summarise(n0 = n())
cbind(gender=table_gender$KEY_SEX_1997,
      share_of_0=table_gender0$n0/table_gender$n)
```

```
##      gender  share_of_0
## [1,]      1 0.007500000
## [2,]      2 0.005742726
```

```r
# iii) by number of children and marital status
# children
table_child = dat132 %>%
  filter(!is.na(CV_BIO_CHILD_HH_U18_2019)) %>%
  group_by(CV_BIO_CHILD_HH_U18_2019) %>%
  summarise(n = n())
table_child0 = dat132_0 %>%
  filter(!is.na(CV_BIO_CHILD_HH_U18_2019)) %>%
  group_by(CV_BIO_CHILD_HH_U18_2019) %>%
  summarise(n0 = n())
table_child0 = rbind(table_child0,data.frame(CV_BIO_CHILD_HH_U18_2019=4:9,n0=c(0)*6))
cbind(children=table_child$CV_BIO_CHILD_HH_U18_2019,
      share_of_0=table_child0$n0/table_child$n)
```

```
##       children  share_of_0
##  [1,]        0 0.014897579
##  [2,]        1 0.007846556
##  [3,]        2 0.005743001
##  [4,]        3 0.008025682
##  [5,]        4 0.000000000
##  [6,]        5 0.000000000
##  [7,]        6 0.000000000
##  [8,]        7 0.000000000
##  [9,]        8 0.000000000
## [10,]        9 0.000000000
```

```r
# marital status
table_marital = dat132 %>%
  filter(!is.na(CV_MARSTAT_COLLAPSED_2019)) %>%
  group_by(CV_MARSTAT_COLLAPSED_2019) %>%
  summarise(n = n())
table_marital0 = dat132_0 %>%
  filter(!is.na(CV_MARSTAT_COLLAPSED_2019)) %>%
  group_by(CV_MARSTAT_COLLAPSED_2019) %>%
  summarise(n0 = n())
table_marital0 = rbind(table_marital0,data.frame(CV_MARSTAT_COLLAPSED_2019=4,n0=0))
cbind(marital_status=table_marital$CV_MARSTAT_COLLAPSED_2019,
      share_of_0=table_marital0$n0/table_marital$n)
```

```
##      marital_status  share_of_0
## [1,]              0 0.005649718
## [2,]              1 0.007454342
## [3,]              2 0.043010753
## [4,]              3 0.001538462
## [5,]              4 0.000000000
```

1.3.3 Interpret the visualization from above

In total, we have 5412 non-missing observations for the income variable, 36 of them have zero income. So on average, the share of "0" in the income data for the whole population is 0.006651885. For different age groups, people with age 35 are more likely to have zero income than the other age groups. They also have relatively low positive incomes. Although males are more likely to have zero income, males who have positive incomes earn more than females. People having no child or being divorced are most likely to have zero income.

Exercise 2 Heckman Selection Model

2.1 Specify and estimate an OLS model to explain the income variable (where income is positive).

```
dat['positive_income'] = as.numeric(dat$YINC_1700_2019>0)
dat1 = dat %>%
  filter(!is.na(positive_income),
         !is.na(age),
         !is.na(work_exp),
         !is.na(edu),
         !is.na(KEY_SEX_1997),
         !is.na(CV_BIO_CHILD_HH_U18_2019)) %>%
  mutate(KEY_SEX_1997=KEY_SEX_1997-1,
         work_exp2=work_exp^2)
dat1_positive = dat1[which(dat1$positive_income==1),]
linear_model = lm(YINC_1700_2019~work_exp+work_exp2+edu+KEY_SEX_1997+CV_BIO_CHILD_HH_U18_2019,
                  data=dat1_positive)
```

Table 1: OLS model

	Dependent variable:
	YINC_1700_2019
work_exp	$3,001.4260^{***}$
work_exp2	-100.0394^{***}
edu	$2,069.8960^{***}$
KEY_SEX_1997	$-19,898.5100^{***}$
CV_BIO_CHILD_HH_U18_2019	$1,024.2250^{***}$
Constant	$18,780.2800^{***}$
Observations	3,914
Adjusted R^2	0.3082
Note:	*p<0.1; **p<0.05; ***p<0.01

2.1.1 Interpret the estimation results

Having one more year of schooling increases the yearly income by $2069.8960. Work experience has a quadratic effect on income. As the increase of work experience, people's income would first increase and then decrease. On average, females earn $19898.5100 less than males. In addition, having one more children increases the income by $1024.2250. The effects of education, work experience, gender and number of children are all statistically significant at 1% significant level.

2.1.2 Explain why there might be a selection problem when estimating an OLS this way

There might be a selection problem because although researchers can observe the actual income of employed people, they cannot observe the reservation income of unemployed people with actual income zero.

2.2 Explain why the Heckman selection model can deal with the selection problem.

In the first stage of the Heckman selection model, we use probit model to predict the probability of having positive incomes, namely the inverse Mill ratio. Including the inverse Mill ratio in the second stage equation can yield consistent estimates of other independent variables and obtain the selection effect.

2.3 Estimate a Heckman selection model. Interpret the results from the Heckman selection model and compare the results to OLS results. Why does there exist a difference?

```
# Stage 1: Probit model
probit_stage1 = glm(positive_income~work_exp+work_exp2+edu+KEY_SEX_1997+CV_BIO_CHILD_HH_U18_2019,
             data=dat1,family=binomial(link='probit'))
mill0 = dnorm(predict(probit_stage1))/pnorm(predict(probit_stage1))
imr = mill0[dat1$positive_income==1]

# Stage 2: OLS model
linear_stage2 = lm(YINC_1700_2019~work_exp+work_exp2+edu+KEY_SEX_1997+CV_BIO_CHILD_HH_U18_2019+imr,
               data=dat1_positive)
```

Table 2: OLS model and Heckman selection model

	Dependent variable:		
	positive_income	YINC_1700_2019	
	probit	*OLS*	
	2.2 stage 1	2.2 stage 2	2.1
work_exp	0.0343	$1,559.4130^{***}$	$3,001.4260^{***}$
work_exp2	-0.0007	-66.1974^{***}	-100.0394^{***}
edu	-0.0019	$2,104.9010^{***}$	$2,069.8960^{***}$
KEY_SEX_1997	-0.0247	$-19,353.6000^{***}$	$-19,898.5100^{***}$
CV_BIO_CHILD_HH_U18_2019	0.1200^{*}	$-3,203.4200^{***}$	$1,024.2250^{***}$
imr		$-755,473.8000^{***}$	
Constant	2.1003^{***}	$48,790.6700^{***}$	$18,780.2800^{***}$
Observations	3,944	3,914	3,914
Adjusted R^2		0.3108	0.3082
Note:			*p<0.1; **p<0.05; ***p<0.01

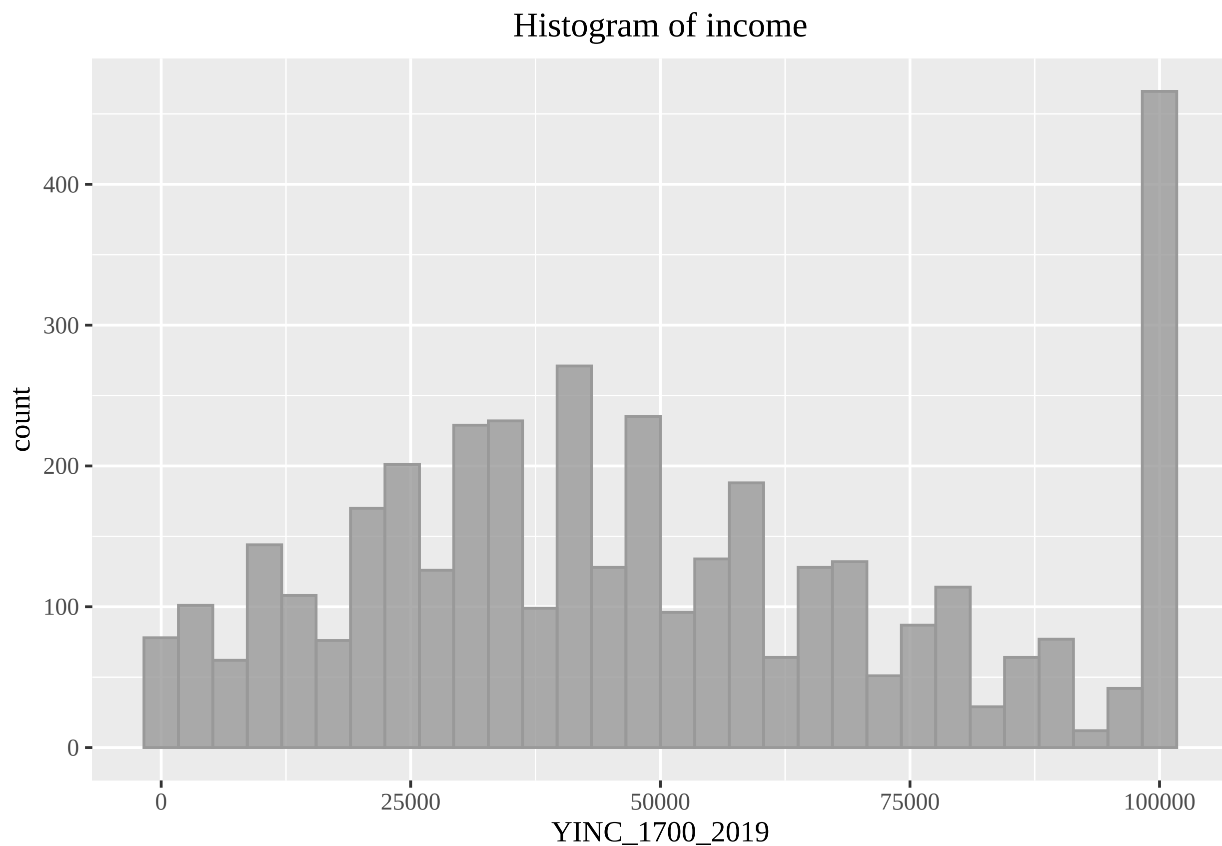
From the second stage of Heckman selection model, we find that having one more year of schooling increases the yearly income by $2104.9010. Work experience still has a quadratic effect on income. On average, females earn $19353.6000 less than males.The different thing is, having one more children would decrease the income by $3203.4200 rather than increase. The selection effect/effect of imr is statistically significant at 1% significant level, which means the dataset indeed suffer from a selection problem. The existence of the difference is due to the neglect of reservation incomes for people with zero income. After taking reservation incomes into consideration, the slope of main factor (i.e. `work_exp`) should be flatter than that in **2.1** OLS model.

Exercise 3 Censoring

Note that the YINC-1700 variable is censored because of privacy issues. In other words, high wages are top-coded in this dataset.

3.1 Plot a histogram to check whether the distribution of the income variable. What might be the censored value here?

```
ggplot(dat1,aes(x=YINC_1700_2019)) +
  geom_histogram(color='#999999', fill='#999999',alpha=0.8) +
  ggtitle('Histogram of income') +
  theme(plot.title = element_text(hjust = 0.5),text=element_text(family='Times'))
```



The censored value here is 100,000.

3.2 Propose a model to deal with the censoring problem.

We should use the censored regression model, which is a generalization of the standard **Tobit** model, to deal with the censoring problem.

3.3 Estimate the appropriate model with the censored data.

```
# the likelihood function
Flike = function(B,x1,x2,x3,x4,x5,y){
  a = -Inf
  b = 100000 # censored value
  Xbeta = B[1]+B[2]*x1+B[3]*x2+B[4]*x3+B[5]*x4+B[6]*x5
  s = B[7] # sigma
```

```r
  cdf1 = pnorm((a-Xbeta)/s)
  cdf2 = pnorm((Xbeta-b)/s)
  pdf_z = dnorm((y-Xbeta)/s)
  cdf1[cdf1>0.999999] = 0.999999
  cdf1[cdf1<0.000001] = 0.000001
  cdf2[cdf2>0.999999] = 0.999999
  cdf2[cdf2<0.000001] = 0.000001
  pdf_z[pdf_z<0.000001] = 0.000001

  I_a = as.numeric(y==a)
  I_b = as.numeric(y==b)
  log_likelihood = sum(I_a*log(cdf1) + I_b*log(cdf2) + (1-I_a-I_b)*(log(pdf_z)-log(s)))
  return(-log_likelihood)
}

# optimization
num_try = 10
out = mat.or.vec(num_try,8)
for (i in 1:num_try) {
  set.seed(i)
  start = c(runif(1,17000,17050),
            runif(1,3200,3250),
            runif(1,-110,-105),
            runif(1,2200,2250),
            runif(1,-22000,-21500),
            runif(1,1300,1400),
            runif(1,26500,27000))
  res = optim(start,fn=Flike,method='BFGS',control=list(trace=6,maxit=1000),
              x1=dat1$work_exp,
              x2=dat1$work_exp2,
              x3=dat1$edu,
              x4=dat1$KEY_SEX_1997,
              x5=dat1$CV_BIO_CHILD_HH_U18_2019,
              y=dat1$YINC_1700_2019)
  out[i,] = c(res$par,res$value)
}
```

```
## initial  value 40957.185341
## final  value 40957.154899
## converged
## initial  value 40957.707465
## final  value 40957.661910
## converged
## initial  value 40957.775436
## final  value 40957.626604
## converged
## initial  value 40958.772249
## final  value 40957.098567
## converged
## initial  value 40957.606450
## final  value 40957.046287
## converged
## initial  value 40957.365361
```

```
## final  value 40957.273661
## converged
## initial  value 40959.215521
## final  value 40957.141948
## converged
## initial  value 40957.390974
## final  value 40957.364303
## converged
## initial  value 40959.746231
## final  value 40957.139086
## converged
## initial  value 40957.901325
## final  value 40957.520465
## converged
```

```
out[which(out[,8]==min(out[,8]))[1],]
```

```
## [1]  17012.5923   3244.1029   -108.1609   2252.2110 -21946.0502   1374.9006
## [7]  26764.7071  40957.0463
```

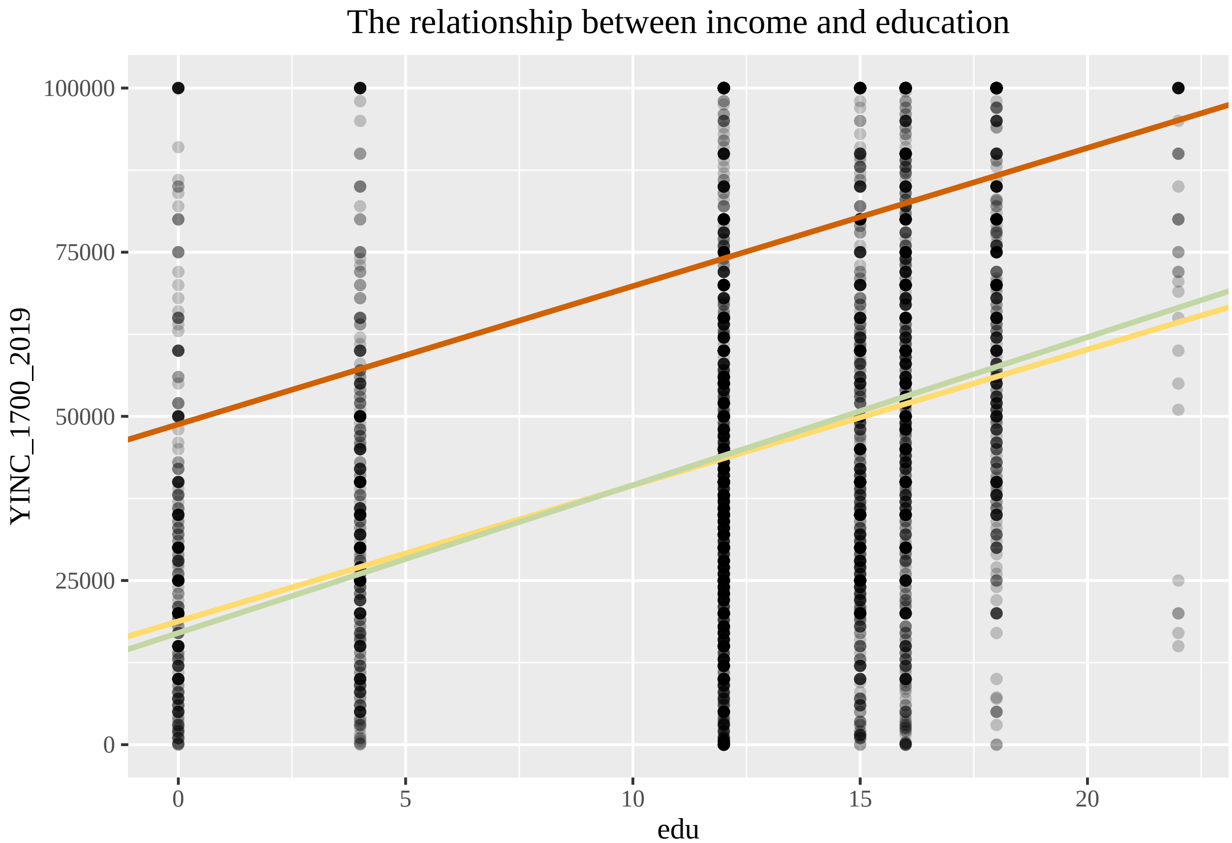
3.4 Interpret the results above and compare to those when not correcting for the censored data.

```
cbind('2.1 OLS'=linear_model$coefficients,
      '2.2 Heckman'=linear_stage2$coefficients[-7],
      '2.3 Tobit'=out[which(out[,8]==min(out[,8]))[1],c(-7,-8)])
```

```
##                              2.1 OLS  2.2 Heckman   2.3 Tobit
## (Intercept)               18780.2799  48790.67298  17012.5923
## work_exp                   3001.4259   1559.41341   3244.1029
## work_exp2                  -100.0394    -66.19741   -108.1609
## edu                        2069.8955   2104.90066   2252.2110
## KEY_SEX_1997             -19898.5091 -19353.60306 -21946.0502
## CV_BIO_CHILD_HH_U18_2019   1024.2250  -3203.42005   1374.9006
```

From the Tobit model, we find that having one more year of schooling increases the yearly income by nearly \$2250. Work experience has a larger quadratic effect on income than the OLS in 2.1. On average, females earn nearly \$22000 less than males. The effect of number of children is also positive.

Above all, it is easy to notice that Tobit model has the steepest slopes for all variables, whereas Heckman selection model has the flatter slopes. Take the slope of `edu` as an example, following figure shows the relationship between income and education. The red line represents the Heckman selection model, the yellow line represents the OLS model, and the green line represents the Tobit model.



Exercise 4 Panel Data

```
dat_panel_raw = read.csv('dat_A4_panel.csv', header=TRUE)
dat4 = dat_panel_raw

# Data pre-processing
# work experience
years = c(1997:2010,seq(2011,2019,2))
for (t in years) {
  wkswk_t = names(dat4)[grepl('WKSWK',names(dat4)) & grepl(as.character(t),names(dat4))]
  dat4 = dat4 %>%
    mutate_at(wkswk_t,funs(ifelse(is.na(.),0,.)))
  dat4[paste0('work_exp_',t)] = rowSums(dat4[,wkswk_t])/52
}

# education
# if edu is NA,fill it with previous year's education
edu_vnames = names(dat4)[grepl('DEGREE',names(dat4))]
edu_vnames = edu_vnames[c(-16,-18)]
dat4 = dat4 %>%
  mutate_at('CV_HIGHEST_DEGREE_9899_1998',funs(ifelse(is.na(.),0,.)))
for (i in 1:dim(dat4)[1]) {
  for (t in 2:19) {
    edu_t_minus1 = edu_vnames[t-1]
```

```r
    edu_t = edu_vnames[t]
    if (is.na(dat4[i,edu_t])){
      dat4[i,edu_t] = dat4[i,edu_t_minus1]
    }
  }
}

# marital status
marital_vnames = names(dat4)[grepl('MARSTAT',names(dat4))]

# wages
income_vnames = names(dat4)[grepl('YINC',names(dat4))]
```

```r
# convert data from Wide to Long
dat_wide = dat4[,c(edu_vnames,marital_vnames,
                   paste0('work_exp_',years),income_vnames)]
names(dat_wide) = c(paste0('edu_',years),paste0('marital_',years),
                    paste0('work_exp_',years),paste0('income_',years))
dat_long = long_panel(dat_wide,prefix='_',begin=1997,end=2019,label_location='end')
dat_long = dat_long %>%
  filter(wave!=2012,wave!=2014,wave!=2016,wave!=2018)
dat_long4 = dat_long %>%
  filter(!is.na(work_exp),!is.na(marital),!is.na(income))
dat_long4 = dat_long4 %>%
  group_by(id) %>%
  mutate(panel_length=n())
dat_long4['edu_year'] = case_when(
  dat_long4$edu==0 ~ 0,  # None
  dat_long4$edu==1 ~ 4,  # GED
  dat_long4$edu==2 ~ 12, # High school diploma
  dat_long4$edu==3 ~ 15, # AA
  dat_long4$edu==4 ~ 16, # BA, BS
  dat_long4$edu==5 ~ 18, # MA, MS
  dat_long4$edu==6 ~ 22, # PhD
  dat_long4$edu==7 ~ 18  # DDS, JD, MD
)
dat_long4['marital_NM'] = as.numeric(dat_long4$marital==0) # Never married
dat_long4['marital_M'] = as.numeric(dat_long4$marital==1) # Married
dat_long4['marital_S'] = as.numeric(dat_long4$marital==2) # Separated
dat_long4['marital_D'] = as.numeric(dat_long4$marital==3) # Divorced
dat_long4['marital_W'] = as.numeric(dat_long4$marital==4) # Widowed

# average x and y over time
dat_long4 = dat_long4 %>%
  group_by(id) %>%
  mutate(avg_income=mean(income),
         avg_edu_year=mean(edu_year),
         avg_work_exp=mean(work_exp),
         avg_marital_NM=mean(marital_NM),
         avg_marital_M=mean(marital_M),
         avg_marital_S=mean(marital_S),
         avg_marital_D=mean(marital_D),
         avg_marital_W=mean(marital_W),
```

```
        )
# xi-x_avg, yi-y_avg
dat_long4 = dat_long4 %>%
  mutate(income_minus_avg=income-avg_income,
         edu_year_minus_avg=edu_year-avg_edu_year,
         work_exp_minus_avg=work_exp-avg_work_exp,
         marital_NM_minus_avg=marital_NM-avg_marital_NM,
         marital_M_minus_avg=marital_M-avg_marital_M,
         marital_S_minus_avg=marital_S-avg_marital_S,
         marital_D_minus_avg=marital_D-avg_marital_D,
         marital_W_minus_avg=marital_W-avg_marital_W)

# x_(t)-x_(t-1), y_(t)-y_(t-1)
dat_long4 = dat_long4 %>%
  group_by(id) %>%
  mutate(diff_income=income-dplyr::lag(income,n=1,default=NA),
         diff_edu_year=edu_year-dplyr::lag(edu_year,n=1,default=NA),
         diff_work_exp=work_exp-dplyr::lag(work_exp,n=1,default=NA),
         diff_marital_NM=marital_NM-dplyr::lag(marital_NM,n=1,default=NA),
         diff_marital_M=marital_M-dplyr::lag(marital_M,n=1,default=NA),
         diff_marital_S=edu_year-dplyr::lag(marital_S,n=1,default=NA),
         diff_marital_D=edu_year-dplyr::lag(marital_D,n=1,default=NA),
         diff_marital_W=edu_year-dplyr::lag(marital_W,n=1,default=NA))
d = data.frame(dat_long4)
```

4.1 Explain the potential ability bias when trying to explain to understand the determinants of wages

Some people are inherently more able than others (ability includes in error term), and these people tend to get higher education (x) and higher wages (y). The error term is correlated with both x and y, so the potential ability bias arises.

4.2 Exploit the panel dimension of the data to propose a model to correct for the ability bias. Estimate the model using the following strategy.

4.2.1 Within Estimator

$$y_{i,t} - \overline{y_i} = \beta(x_{i,t} - \overline{x_i}) + (\epsilon_{i,t} - \overline{\epsilon_i})$$

```
formula_within = income_minus_avg~edu_year_minus_avg+work_exp_minus_avg+
  marital_M_minus_avg+marital_S_minus_avg+marital_D_minus_avg+
  marital_W_minus_avg-1
within_model = lm(formula_within,data=dat_long4)
```

4.2.2 Between Estimator

$$\overline{y_i} = \alpha_i - \beta \cdot \overline{x_i} + \overline{\epsilon_i}$$

```
btw_data = d[,c('id','avg_income','avg_edu_year','avg_work_exp',
                'avg_marital_M','avg_marital_S','avg_marital_D',
                'avg_marital_W')]
btw_data = distinct(btw_data,id,.keep_all = TRUE)

formula_btw = avg_income~avg_edu_year+avg_work_exp+
  avg_marital_M+avg_marital_S+avg_marital_D+avg_marital_W
btw_model = lm(formula_btw,data=btw_data)
```

4.2.3 Difference (any) Estimator

$$y_{i,t} - y_{i,t-1} = \beta(x_{i,t} - x_{i,t-1}) + (\epsilon_{i,t} - \epsilon_{i,t-1})$$

```
formula_fd = diff_income~diff_edu_year+diff_work_exp+
  diff_marital_M+diff_marital_M+diff_marital_S+diff_marital_D+
  diff_marital_W-1
fd_model = lm(formula_fd,data=dat_long4[!is.na(d$diff_income),])
```

Table 3: Model estimation using three strategies

	Dependent variable:		
	income_minus_avg Within	avg_income Between	diff_income First difference
edu_year_minus_avg	$2,019.9450^{***}$		
work_exp_minus_avg	$2,631.2400^{***}$		
marital_M_minus_avg	$19,803.2200^{***}$		
marital_S_minus_avg	$16,327.6000^{***}$		
marital_D_minus_avg	$20,381.5800^{***}$		
marital_W_minus_avg	$11,068.0500^{***}$		
avg_edu_year		876.6572^{***}	
avg_work_exp		$2,178.7660^{***}$	
avg_marital_M		$9,147.1500^{***}$	
avg_marital_S		706.4367	
avg_marital_D		$2,831.9620^{**}$	
avg_marital_W		$-19,715.4300^{**}$	
diff_edu_year			97.6767^{***}
diff_work_exp			956.6376^{***}
diff_marital_M			$2,313.6340^{***}$
diff_marital_S			-258.0530
diff_marital_D			$1,103.4820^{***}$
diff_marital_W			-502.5369
Constant		$4,660.7850^{***}$	
Observations	83,306	8,637	74,669
Adjusted R^2	0.3336	0.2159	0.0915
Note:			*p<0.1; **p<0.05; ***p<0.01

4.3 Interpret the results from each model and explain why different models yield different parameter estimates

In short, different models yield different parameter estimates because they measure different things.

Both within estimator (FE) and first difference (FD) estimator measure variation of individual over time, all time-invariant variables (i.e. ability) including constant for each individual i drop out of the model.

For FE, each variable is demeaned and FE is preferred when the processes are weakly dependent over time. When a person has one more year of education than her yearly average, her income would increase by \$2019.9450; similarly, when a person has one more year of work experience than her yearly average, her income would increase by \$2631.2400.

For FD, each variable is differenced once over time, so we can estimate the relationship between changes of variables. It has advantage for processes with large positive autocorrelation. If a person obtain one more year of education or work experience than previous year, her wage is predicted to increase by \$97.6767 or \$956.6376.

Between estimator measures variation of the means across individuals, we estimate α and β using the cross-sectional information in the data. We find that if person's average years of education increases one year, her wage would increase by \$876.7572; if her work experience increases one year, her wage would increase by \$2178.7660.