

יישומים אקונומטריים - תרגיל 3

$$Var(u_i) = \sigma^2 r_i^2 s_i^2$$

א. ל.

החוקר יודע שהקטעים שונים מהסדר $Var(u_i)$ ולכן יכול
לשערר על בגודל WLS ולכן לקבל אומרים
לקובעים אסטרטגיה.

$$Var(u_i | x_i) = \sigma^2 h_i = \sigma^2 r_i^2 s_i^2$$

בהקשר של $h_i = r_i^2 s_i^2$

נאמר - המעבר ב- $r_i s_i = \sqrt{h_i}$

$$\frac{y_i}{r_i s_i} = \frac{\beta_0}{r_i s_i} + \frac{\beta_1 q_i}{r_i s_i} + \frac{\beta_2 \cancel{x_i}}{\cancel{r_i s_i}} + \frac{\beta_3 r_i^2}{r_i s_i} + \frac{\beta_4 \cancel{s_i}}{\cancel{r_i s_i}} + u_i$$

$$\frac{y_i}{r_i s_i} = \frac{\beta_0}{r_i s_i} + \frac{\beta_1 q_i}{r_i s_i} + \frac{\beta_2}{s_i} + \frac{\beta_3 r_i}{s_i} + \frac{\beta_4}{r_i} + u_i$$

משתנה טרנספורמ

משתנים טרנספורמ

כל המחקר נעשה זה השווה - בצורה אחידה בין כל המשתנים
לכן יצר שווה - אחידה, ככה מקיימים כל הנתונים וקטגוריות
וקנה למשל מהאומרים (באופן) מספר המים והקטגוריות, אומרים סוג
וערך מספר המים והקטגוריות וכלי שווה - מניחים.

כאובן של המים שמים ב- WLS האומרים שהם קטגוריות (באופן) וואריאנסיות

שונים המים שהם מחרמים ב- $\sqrt{h_i}$ ובצורה משוואה קיבלו

אומרים שהם ידועים והם מקיים ב- h_i הפונקציה הקטגוריות.

חברו לכן ששונה המים למחרם עסי ושלם:

$$\begin{aligned} Var(u_i^* | x_i^*) &= E(u_i^{*2} | x_i^*) = E\left(\frac{u_i}{\sqrt{h_i}}\right)^2 | x_i^* = E\left(\frac{u_i^2}{h_i} | x_i^*\right) = \\ &= \frac{1}{h_i} \cdot E(u_i^2 | x_i^*) = \frac{1}{h_i} Var(u_i | x_i^*) = \frac{1}{h_i} Var(u_i | x_i) = \frac{1}{h_i} \sigma^2 h_i = \sigma^2 \end{aligned}$$

למשל שווה בגודל המים.

ב. במידה והפרק יורג רק שמה קיים $Var(u_i) = \sigma^2$ כלומר אינו יוצא באצט צורה השונה האוסטרקסטי, יהיה טליו למשל בתיקון White על מנת לקבל אומנים עקבים לסל-תן החוקר ואומנו זל-סטיו-התקן לפי הנוסחה:

$$Var(\hat{\beta}_j) = \frac{\sum_{i=1}^n \hat{u}_i^2}{SSR_j}$$

גם לפי שיטה זו עוקמים א-הסיכור חוזק שנתפז סאר המשתנים ומפצרים אומן בין סאר המשתנים במאז. לפי שיטה זו אין צורך לטול בפיקן מהו מבקד השיוון אך היא לא מספקת אומנים לריאליזם (בלאט) יחיליה (BLA) כמי השיטה WLS.

האומנים לפימאליז לא משגים ובקז R^2 לא משגים, אומני סטיו-התקן כן משגים והאומנים לא יחיו יחיליה (שאר טליו המושגית לא השגו).

2. א. נקרא את קובץ ה-CSV ונגביל אותו לגילאים המבוקשים שדיווחו על שנות לימוד, נקבל מסד נתונים עם 9,558 תצפיות ו-10 משתנים, נשים לב שחסרות תצפיות עבור השכר.

```
> #Import Data
> ps3=read.csv("ps3.csv")
> #check "schooling" var
> # limit the data for men in ages 25-66
> # with reported education only
> ps3_new = ps3[ps3$age <= 66 & ps3$age >= 25 & !is.na(ps3$schooling), ]
> describe(ps3_new)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
age	1	9558	43.61	12.08	42.00	43.31	14.83	25.00	66.00	41.00	0.17	-1.18	0.12
arab	2	9558	0.16	0.36	0.00	0.07	0.00	0.00	1.00	1.00	1.88	1.54	0.00
children_0_4	3	9558	0.41	0.74	0.00	0.24	0.00	0.00	4.00	4.00	1.78	2.55	0.01
lwage_2011	4	7262	3.83	0.64	3.74	3.80	0.64	-0.83	7.25	8.08	0.26	1.48	0.01
marital	5	9558	1.64	1.20	1.00	1.41	0.00	1.00	5.00	4.00	1.47	0.31	0.01
schooling	6	9558	13.79	4.24	13.00	13.65	2.97	0.00	59.00	59.00	1.31	9.32	0.04
status	7	9558	1.73	1.28	1.00	1.54	0.00	1.00	4.00	3.00	1.20	-0.55	0.01
yeshiva	8	9558	0.05	0.22	0.00	0.00	0.00	0.00	1.00	1.00	4.06	14.48	0.00

ב. נגדיר משתנה חדש לסלקציה עבור האוכלוסיה המשתכרת :

```
> # Q2
> # new var for selection if the obs is salaried worker
> ps3_new$selection = as.numeric(ps3_new$status==1|ps3_new$status== 2 )
> table(ps3_new$selection, useNA = 'always')
```

	0	1	<NA>
	2296	7262	0

ג. נוסף משתנה "גיל בריבוע" ונאמוד משוואת Mincer לשכר:

$$\text{Log}(\text{wage}) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \beta_3 \text{Schooling} + \beta_4 \text{Arab} + u$$


```

> # Q3
> ps3_new$agesq = (ps3_new$age)^2
> mincerModel<-lm(lwage_2011 ~ age + agesq + schooling + arab, ps3_new)
> options(scipen=999, digits=4)
> summary(mincerModel)

Call:
lm(formula = lwage_2011 ~ age + agesq + schooling + arab, data = ps3_new)

Residuals:
    Min       1Q   Median       3Q      Max
-4.686 -0.341 -0.026  0.343  2.981

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.2585976   0.1028455   12.2 <0.0000000000000002 ***
age          0.0697608   0.0047349   14.7 <0.0000000000000002 ***
agesq       -0.0006769   0.0000532  -12.7 <0.0000000000000002 ***
schooling    0.0679234   0.0018907   35.9 <0.0000000000000002 ***
arab        -0.2491333   0.0197535  -12.6 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.557 on 7257 degrees of freedom
(2296 observations deleted due to missingness)
Multiple R-squared:  0.242,    Adjusted R-squared:  0.242
F-statistic: 580 on 4 and 7257 DF,  p-value: <0.0000000000000002

```

- המקדם של schooling, כלומר התשואה להשכלה היא 0.068 בקירוב.

$$\beta_3 = 0.0679234$$

- יש מקום לחשוד שהאומד מוטה משום ש- 2296 תצפיות, כפי שמצוין בפלט, הוסרו עקב היעדר נתונים לגבי השכר, כלומר לא ניתן לחזות לגביהן משום שתצפיות אלה לא דיווחו שכר, מה שמעיד על חשש להטיית סלקטיבית.

ד. נכתוב מודל עם תיקון לסלקציה בשיטת Heckman בשני שלבים:

שלב #1 – נאמוד מודל *Probit* למשתנה סלקציה (*S*) כמשתנה מוסבר ע"י משתני *Z* שיכילו את כלל המשתנים המסבירים ממודל OLS בתוספת משתנה דמי-ילדים בגיל הרך:

$$Selected\ Workers = \gamma_0 + \gamma_1 Age + \gamma_2 Age^2 + \gamma_3 Schooling + \gamma_4 Arab + \gamma_5 children_{0_4} + v$$

שלב #2 – נריץ רגרסיה ליניארית רק על המדגם החלקי (אלו שעובדים):

$$\log(wage) = \beta_0 + \beta_1 Age + \beta_2 Age^2 + \beta_3 Schooling + \beta_4 Arab + u$$

- על מנת לקבל אומדים עקיבים נצטרך את התנאים הבאים:

$$cov(u|z) = 0$$

$$cov(v|z) = 0$$

$$cov(v|x) = 0$$

$$cov(u|x) = 0$$

- ובנוסף נדרוש שההפרעה במודל הסלקציה תתפלג נורמלית:

$$v \sim N(\mu, \sigma)$$

ה. נריץ את המודל ונקבל:

```
> #performs a 2-step Heckman (heckit) estimation that corrects for
non-random sample selection.
> heckmanModel <- heckit(selection ~ age + agesq + schooling + arab
+ children_0_4 ,lwage_2011 ~ age + agesq + schooling + arab, ps3_ne
w)
> summary(heckmanModel)
-----
Tobit 2 model (sample selection model)
2-step Heckman / heckit estimation
9558 observations (2296 censored and 7262 observed)
14 free parameters (df = 9545)
Probit selection equation:
      Estimate Std. Error t value      Pr(>|t|)
(Intercept) -2.071599   0.211452  -9.80 < 0.0000000000000002 ***
age          0.148079   0.009806  15.10 < 0.0000000000000002 ***
agesq       -0.001829   0.000109  -16.74 < 0.0000000000000002 ***
schooling    0.014466   0.003373   4.29   0.0000181677 ***
arab        -0.340648   0.039804  -8.56 < 0.0000000000000002 ***
children_0_4 -0.120837   0.020683  -5.84   0.0000000053 ***
Outcome equation:
      Estimate Std. Error t value      Pr(>|t|)
(Intercept)  0.358209   0.330170   1.08   0.28
age          0.105289   0.013321   7.90   0.0000000000000003 ***
agesq       -0.001112   0.000161  -6.90   0.0000000000005442 ***
schooling    0.070788   0.002264  31.27 < 0.0000000000000002 ***
arab        -0.333747   0.036333  -9.19 < 0.0000000000000002 ***
Multiple R-Squared:0.243,      Adjusted R-Squared:0.243
Error terms:
      Estimate Std. Error t value Pr(>|t|)
invMillsRatio  0.533      0.182   2.92  0.0035 **
sigma          0.658      NA      NA      NA
rho            0.810      NA      NA      NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- נקבל $\hat{\beta}$ שהוא עקיב ומתפלג נורמלית בקירוב.

שלב #3 – נבדוק הטיית סלקציה ע"י ההשערה:

$$H_0: \hat{\beta}_\lambda = 0$$

$$H_1: \hat{\beta}_\lambda \neq 0$$

- נביט על אומד יחס מילס ההופכי ($\hat{\beta}_\lambda$ -נמצא מובהק מאוד) שנמצא בפלט וניווכח כי:

$$\hat{\beta}_\lambda = 0.533$$

- לכן נדחה את השערת H_0 ונגיד כי קיימת הטעייה של האומד ברמת בטחון של 99%.
- נבדוק את ערך β_3 של Schooling – האומד לתשואה להשכלה לפי מודל OLS ונשווה אותו לערך האומד לפי מודל Heckman ונמצא שהאומד לפי Heckman גבוהה יותר כלומר מצאנו הטייה כלפי מטה במודל הלינארי.

$$\beta_{3 \text{ Heckman}} = 0.0707 > 0.0679 = \beta_{3 \text{ OLS}}$$

1. נחזור על הסעיפים ג' ו-ה' רק עבור גברים נשואים, נקבל את אמידת המודל Mincer הבא:

```
> # Q2.E
> mincerModel_men<-lm(lwage_2011 ~ age+ agesq+ schooling+ arab, p
s3_new[ps3_new$marital ==1, ])
> summary(mincerModel_men)

Call:
lm(formula = lwage_2011 ~ age + agesq + schooling + arab, data =
ps3_new[ps3_new$marital ==
1, ])

Residuals:
    Min       1Q   Median       3Q      Max
-4.715 -0.349 -0.030  0.347  2.980

Coefficients:
            Estimate Std. Error t value      Pr(>|t|)
(Intercept)  1.4917765   0.1401403   10.64 <0.0000000000000002 ***
age           0.0640200   0.0062037    10.32 <0.0000000000000002 ***
agesq        -0.0006328   0.0000674    -9.39 <0.0000000000000002 ***
schooling     0.0656556   0.0020508    32.01 <0.0000000000000002 ***
arab         -0.3026368   0.0221515   -13.66 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.556 on 5628 degrees of freedom
(1594 observations deleted due to missingness)
Multiple R-squared:  0.237,    Adjusted R-squared:  0.237
F-statistic: 437 on 4 and 5628 DF, p-value: <0.0000000000000002
```

- נוכל לראות כי גם הפעם הוסרו 1594 תצפיות חסרות כלומר יש חשד להטיית סלקציה, לכן נתקן באמצעות שיטת 2 השלבים של Heckman ונקבל:

```
> # 2 step correction only for men
> heckman_men<- heckit(selection ~age + agesq + schooling +arab +
children_0_4, lwage_2011 ~ age +agesq +schooling +arab , ps3_new[p
s3_new$marital ==1, ])
> summary(heckman_men)

-----
Tobit 2 model (sample selection model)
2-step Heckman / heckit estimation
7227 observations (1594 censored and 5633 observed)
14 free parameters (df = 7214)
Probit selection equation:
            Estimate Std. Error t value      Pr(>|t|)
(Intercept) -1.248002   0.319651   -3.90    0.000095 ***
age           0.139461   0.013681   10.19 < 0.0000000000000002 ***
agesq        -0.001841   0.000144   -12.78 < 0.0000000000000002 ***
schooling     0.000502   0.003728    0.13    0.89
arab         -0.508437   0.046200   -11.01 < 0.0000000000000002 ***
children_0_4 -0.262555   0.026412   -9.94 < 0.0000000000000002 ***
Outcome equation:
            Estimate Std. Error t value      Pr(>|t|)
(Intercept)  1.591657   0.226077    7.04    0.000000000000021 ***
age           0.059670   0.009908    6.02    0.0000000018022 ***
agesq        -0.000580   0.000116   -5.01    0.0000005532471 ***
schooling     0.065729   0.002056   31.97 < 0.0000000000000002 ***
arab         -0.289904   0.031658   -9.16 < 0.0000000000000002 ***
Multiple R-Squared:0.237,    Adjusted R-Squared:0.237
Error terms:
            Estimate Std. Error t value Pr(>|t|)
invMillsRatio -0.058    0.103   -0.56    0.57
sigma          0.557         NA      NA      NA
rho           -0.104         NA      NA      NA
-----
```

$$H_0: \hat{\beta}_{\lambda_{\text{married}}} = 0$$

$$H_1: \hat{\beta}_{\lambda_{\text{married}}} \neq 0$$

$$\hat{\beta}_{\lambda_{\text{married}}} = -0.058$$

- נביט על אומד יחס מילס ההופכי ($\hat{\beta}_{\lambda_{\text{married}}}$) וניווכח כי אינו מובהק כלל, לכן לא נוכל לדחות את השערת H_0 , נבין כי ההתאמה בין ההפרעה u ל- v אינה מובהקת וסביר שהם בלתי מתואמים - $cov(v|u) = 0$, נסיק כי אין הטיית סלקציה כאשר דוגמים את אוכלוסיית הנשואים.

3. /c. המודל יאמר לי מהם המצבים:

$$\text{Average-grade} = \beta_0 + \beta_1 \cdot \text{male} + \beta_2 \cdot \text{University} + \beta_3 \cdot \text{study-field} + \beta_4 \cdot \text{SAT} + \beta_5 \cdot \text{Bagrut-Ave} + \beta_6 \cdot \text{strt-age} + \beta_7 \cdot \text{fath-educ} + \beta_8 \cdot \text{sibil} + \beta_9 \cdot \text{father-inc} + \beta_{10} \cdot \text{region}$$

הנתונים המדוברים: (לדוגמה: המצב של המצב 3.11 - Average-grade)

הנתונים:

male - נקבה 1, גבר 0

University - נקבה 1, גבר 0

study-field - נקבה 1, גבר 0 (הנתונים הם: המצב של המצב 3.11 - Average-grade)

Fath-educ - נקבה 1, גבר 0 (הנתונים הם: המצב של המצב 3.11 - Average-grade)

region - נקבה 1, גבר 0 (הנתונים הם: המצב של המצב 3.11 - Average-grade)

הנתונים (הנתונים הם: המצב של המצב 3.11 - Average-grade):

SAT - המצב של המצב 3.11 - Average-grade

father-inc - המצב של המצב 3.11 - Average-grade

ה. כיווני התקדמות ארוג אני
 משימות חיוביות
 משימות שליליות

Male - סביר שלמים מנחם יגד
 start-age - של מיליון איחוד סמן התחלה בשל
 מאחר יגד משימה שלילית.
 } לאר היתרבות - לפי
 משימה הפה' שלם.

3 יון הסטטיסטי - SAT
 Bagrut-ave - המבחן הסטטיסטי
 father-inc - הכנסה ראש
 Sibi - מספר אחים ואחיות
 (דצבים בלימודים וניסיון דבר)

אונברסיטה בן זוריון (שמה דמי)
 } האוניברסיטה הזרמים

התנהגות
 שגויה מדי
 } התנהגות
 שגויה מדי
 } התנהגות
 שגויה מדי

ז. יש מקום לחשוד בהטיית סלקציה שכן בנהגים קיימים חק
 מיכל אלו - ראשית האלונים בדור שאין לנו במחשבים פסאיד שלם סיימ
 אחרתהי הניסיון לפרש אתהי שובלים גם הם להסביר בצורה טובה לא
 ניסיון הניסיון של הפסדים ולכן יתכן וקרה אחרים מולם
 לפי מיליון דבר הניסיון של אלה פסלים שיכולים להסביר לא ניסיון הניסיון
 ד. לא ניתן למדוד מיליון Heckman במקרה הניסיון מושג שיש לנו
 (תנינים אק ורן דל קול מיליון דל)

במידה והיה לנו תנינים גם דל אוכלוסיות אחרות שיכולים להסביר
 לא מיליון תנינים כמו אלו של סיימ מיליון דל. בשל הניסיון
 הניסיון מיליון דל S (1 = 1: סיימ מיליון דל $S=0$: לא סיימ מיליון דל)
 דל מיליון דל מסקנות במידה והקול יגד מיליון דל (מיליון דל פסלים
 לא מיליון דל $(S=0)$ קרה אחרים $\hat{\theta}$ ונחשב לא יחס מיליון דל
 $prob(s_i=1|z_i) = \phi(z_i, \hat{\theta})$
 בלבד הפני (מיליון דל OLS של y מיליון דל x ו- $\lambda(z_i, \hat{\theta})$
 במידה הניסיון $(S=1) \leftarrow \lambda(z_i, \hat{\theta}) + \beta_2 + \beta_1 x + \epsilon_i$

הניסיון הניסיון שלם הניסיון יתכן הניסיון הניסיון הניסיון
 הניסיון הניסיון הניסיון הניסיון הניסיון הניסיון הניסיון
 הניסיון הניסיון הניסיון הניסיון הניסיון הניסיון הניסיון
 הניסיון הניסיון הניסיון הניסיון הניסיון הניסיון הניסיון

ה. לא ניתן לאתר מודל Heckman משום שהסברים לנו
 נתונים אינם אלא שאלה סיימו תואר ראשון ויש לנו נתונים אך
 ורק של אלו שסיימו הנתונים הנוספים הן מידע אודות סטודנטים
 שאין להם תואר ראשון כלומר שיהיו מדגם מ"צג אינם נוכל לבדוק
 האם ישנה הבדל במידת המאמצים למצוא ← האם סלקציה.

R-SCRIPT

```
install.packages("openxlsx")
install.packages("zoo")
install.packages("lmtest")
install.packages("sampleSelection")
install.packages("visualize")
install.packages('psych')
library("openxlsx")
library('psych')
library(ggplot2)
library(visualize)
library("zoo")
library("lmtest")
library("maxLik")
library("miscTools")
library("sampleSelection")
install.packages("mvtnorm")
library("mvtnorm")
```

Clear Objects

```
rm(list=ls())
```

#removing specific object

```
#rm( x )
```

```
#data$var <- NULL
```

```
#####
```

```
#####
```

Change Directory

```
setwd("/Users/EfiPaka/Desktop/Econometrics/Ex103")
```

#Import Data

```
ps3=read.csv("ps3.csv")
```

Q2

create a dummy var for employment

including absence from work

```
#ps1$emp = as.numeric(ps1$lf_char <= 3)
```

```
#table(ps1$emp, useNA = 'always')
```

Q2.A

#check "schooling" var

limit the data for men in ages 25-66

with reported education only


```
ps3_new = ps3[ps3$age <= 66 & ps3$age >= 25 & !is.na(ps3$schooling) ,]
describe(ps3_new)
```

Q2.B

new var for selection if the obs is salaried worker

```
ps3_new$selection = as.numeric(ps3_new$status== 1 | ps3_new$status== 2 )
table(ps3_new$selection, useNA = 'always')
```

Q2.C

```
ps3_new$agesq = (ps3_new$age)^2
```

```
mincerModel<-lm(lwage_2011 ~ age + agesq + schooling + arab, ps3_new)
```

```
options(scipen=999, digits=4)
```

```
summary(mincerModel)
```

```
jewishSchooling<-lm( ps3_new$lwage_2011[ps3_new$arab ==
0]~ps3_new$schooling[ps3_new$arab == 0])
```

```
summary(jewishSchooling)
```

```
arabSchooling<-lm( ps3_new$lwage_2011[ps3_new$arab ==
1]~ps3_new$schooling[ps3_new$arab == 1])
```

```
summary(arabSchooling)
```

```
describe(ps3_new$arab)
```

```
plot(ps3_new$lwage_2011[ps3_new$arab == 0]~ps3_new$schooling[ps3_new$arab ==
0],main="School Years~Monthly Work Hours", xlab="School Years", ylab="Monthly Wrok
Hours",
```

```
col="blue4", xlim=c(0, 30), ylim=c(0, 8), pch=5, cex=c(0.9,0))
```

```
points(ps3_new$lwage_2011[ps3_new$arab == 1]~ps3_new$schooling[ps3_new$arab ==
1],main="School Years~Monthly Work Hours", xlab="School Years", ylab="Monthly Wrok
Hours",
```

```
col="firebrick1")#, xlim=c(0, 27), ylim=c(0, 600), pch=5, cex=c(0.9,0))
```

```
abline( col="deepskyblue3", lwd=3,coef(jewishSchooling))
```

```
abline( col="firebrick1", lwd=3,coef(arabSchooling))
```

#we want to check how much will i work if i've studyied that much years

#add regression line:

#jewish men

```
#plot(newdata$WPH_In[newdata$national=="jewish" &
newdata$sex=="male"]~newdata$schooly[newdata$national=="jewish" &
newdata$sex=="male"]) ,main="School Years~Log of Wage per Hour- Jewish Men",
xlab="School Years", ylab="Log of Wage per Hour",
```

```
# col="aquamarine4", xlim=c(0, 25), ylim=c(0, 7), pch=11, cex=c(0.9,0)) abline(
col="aquamarine3", lwd=3,coef(modelT3.JM))
```

Q2.D

#performs a 2-step Heckman (heckit) estimation that corrects for non-random sample selection.

```
hecmanModel <- heckit(selection ~ age + agesq + schooling + arab + children_0_4
```

```
,lwage_2011 ~ age + agesq + schooling + arab, ps3_new)
```

```
summary(hecmanModel)
```

Q2.E

```
mincerModel_married<-lm(lwage_2011 ~ age+ agesq+ schooling+ arab,
```

```
ps3_new[ps3_new$marital ==1, ])
```

```
summary(mincerModel_married)
```

```
# 2 step correction only for men
heckman_married<- heckit(selection ~age + agesq + schooling +arab + children_0_4,
lwage_2011 ~ age +agesq +schooling +arab , ps3_new[ps3_new$marital ==1, ])
summary(heckman_married)
```