Section 4: Oaxaca-Blinder Decomposition

Introduction to Econometrics,Fall 2017

Zhaopeng Qu

12/4/2017

Contents

Install Packages

E9.1 Average Hourly Earnings (AHE) and Age

```
cps08 <- read_dta("~/Dropbox/NJU/Teaching/2017Fall/Econometrics/SW_Datasets/cps08.dta")
summary(cps08)
##       ahe              year         bachelor         female
##  Min.   : 2.003   Min.   :2008   Min.   :0.000   Min.   :0.0000
##  1st Qu.:12.019   1st Qu.:2008   1st Qu.:0.000   1st Qu.:0.0000
##  Median :16.827   Median :2008   Median :0.000   Median :0.0000
##  Mean   :18.976   Mean   :2008   Mean   :0.481   Mean   :0.4326
##  3rd Qu.:23.558   3rd Qu.:2008   3rd Qu.:1.000   3rd Qu.:1.0000
##  Max.   :82.418   Max.   :2008   Max.   :1.000   Max.   :1.0000
##       age
##  Min.   :25.00
##  1st Qu.:27.00
##  Median :30.00
##  Mean   :29.58
##  3rd Qu.:32.00
##  Max.   :34.00
```

```
library(sandwich)
library(lfe)
attach(cps08)

fit1 <- lm(ahe ~ age + female + bachelor)
cov1 <- vcovHC(fit1, type = "HC")
robust.se1 <- sqrt(diag(cov1))

# log value of ahe
cps08$lnahe <- log(ahe)

fit2 <- lm(cps08$lnahe ~ age + female + bachelor)
cov2 <- vcovHC(fit2, type = "HC")
robust.se2 <- sqrt(diag(cov2))
```

- age squred

```
cps08$agesq <- age^2
fit3 <- lm(cps08$lnahe ~ age + I(age^2) + female + bachelor)
cov3 <- vcovHC(fit3, type = "HC")
robust.se3 <- sqrt(diag(cov3))
```

- Interactions: female*bachelor and female*age

```
cps08$fxb <- female * bachelor
cps08$fxa <- female * age
cps08$bxa <- bachelor * age
```

```
fit4 <- lm(cps08$lnahe ~ age + I(age^2) + female + bachelor + fxb, data = cps08)
cov4 <- vcovHC(fit4, type = "HC")
robust.se4 <- sqrt(diag(cov4))


fit5 <- lm(lnahe ~ age + I(age^2) + female + bachelor + fxa, data = cps08)
cov5 <- vcovHC(fit5, type = "HC")
robust.se5 <- sqrt(diag(cov5))


fit6 <- lm(cps08$lnahe ~ age + I(age^2) + female + bachelor + bxa, data = cps08)
cov6 <- vcovHC(fit6, type = "HC")
robust.se6 <- sqrt(diag(cov6))
```

- using "stargazer" to produce a publishing-quality table

```
library(stargazer)
stargazer(fit1, fit2, fit3, fit4, fit5, fit6, type = "latex", se = list(robust.se1,
    robust.se2, robust.se3, robust.se4, robust.se5, robust.se6), header = FALSE,
    style = "qje", no.space = TRUE, df = FALSE, notes.align = "l", notes.append = FALSE)
```

Oaxaca-Blinder Decomposition

- Native v.s Foreign

```
library(oaxaca)
data("chicago")
chicago$real.wage <- exp(chicago$ln.real.wage)
results <- oaxaca(formula = real.wage ~ age + female + LTHS + +some.college +
    college + advanced.degree | foreign.born | LTHS + +some.college + college +
    advanced.degree, data = chicago, R = 1000)

results$twofold$overall
##          weight coef(explained) se(explained) coef(unexplained)
## [1,]  0.0000000       1.6165339     0.6902127          1.399040
## [2,]  1.0000000       0.1822482     0.7104507          2.833326
## [3,]  0.5000000       0.8993911     0.5679411          2.116183
## [4,]  0.5690691       0.8003263     0.5724691          2.215248
## [5,] -1.0000000       1.3557222     0.5106738          1.659852
## [6,] -2.0000000       0.9525717     0.5238267          2.063003
##      se(unexplained) coef(unexplained A) se(unexplained A)
## [1,]       0.9368588        1.399040e+00      9.368588e-01
## [2,]       0.8737907        0.000000e+00      0.000000e+00
## [3,]       0.8078351        6.995202e-01      4.684294e-01
```

Table 1:

	ahe		lnahe		lnahe	lnahe
	(1)	(2)	(3)	(4)	(5)	(6)
age	0.585***	0.027***	0.081*	0.081*	0.093**	0.073*
	(0.037)	(0.002)	(0.043)	(0.043)	(0.043)	(0.043)
I(ageˆ2)			−0.001	−0.001	−0.001	−0.001
			(0.001)	(0.001)	(0.001)	(0.001)
female	−3.664***	−0.186***	−0.186***	−0.220***	0.313***	−0.184***
	(0.208)	(0.011)	(0.011)	(0.015)	(0.110)	(0.011)
bachelor	8.083***	0.428***	0.428***	0.398***	0.427***	−0.065
	(0.213)	(0.011)	(0.011)	(0.015)	(0.011)	(0.110)
fxb				0.069***		
				(0.022)		
fxa					−0.017***	
					(0.004)	
bxa						0.017***
						(0.004)
Constant	−0.636	1.876***	1.085*	1.100*	0.803	1.327**
	(1.083)	(0.056)	(0.635)	(0.635)	(0.637)	(0.639)
N	7,711	7,711	7,711	7,711	7,711	7,711
R^2	0.200	0.201	0.201	0.202	0.203	0.203
Adjusted R^2	0.199	0.200	0.200	0.201	0.202	0.202
Residual Std. Error	9.072	0.469	0.469	0.469	0.469	0.469
F Statistic	641.492***	644.876***	484.078***	389.703***	392.217***	392.171***

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

```
## [4,]       0.8049324        6.028898e-01      4.037214e-01
## [5,]       0.6436121        9.445705e-01      3.652813e-01
## [6,]       0.8056315        4.840572e-14      3.964910e-14
##      coef(unexplained B) se(unexplained B)
## [1,]           0.0000000         0.0000000
## [2,]           2.8333261         0.8737907
## [3,]           1.4166630         0.4368954
## [4,]           1.6123582         0.4972473
## [5,]           0.7152816         0.2819896
## [6,]           2.0630026         0.8056315
```

- Plot Figure 1