

Homework 6

Wei Ye*
ECON 7910- Econometrics I

Due on Oct 28, 2021

1 Question 1 – 6.1

Solution:

- a) From the reduced form regression, nearc4 is more correlated with educ, because not only it has higher coefficient, but with higher confidence level. We reject our null hypothesis: the coefficient of nearc4 is 0.

See the table 2, they are completely the same with regarding to coefficients and confidence levels.

- b) From the result of Chi square, it passes the overidentification test.

```
library(tidyverse)
library(stargazer)
library(AER)
library(plm)
#

#6.1 part a
card <- read_csv('card.csv')
head(card, n=5)
iv1 <- ivreg(lwage ~ exper + expersq + black + south + smsa + reg661 + educ | exper + expersq +
            black + south + smsa + reg661 + reg662 + reg663 + reg664 + reg665 + reg666 +
            reg667 + reg668 + nearc2 + nearc4, data=card)
summary(iv1)
olseduc <- lm(educ ~ exper + expersq + black + south + smsa + reg661 + nearc2 + nearc4,
            data=card)
summary(olseduc)
educ_hat <- fitted.values(olseduc)
olslwage <- lm(lwage ~ exper + expersq + black + south + smsa + reg661 + educ_hat, data=card)
```

*1st year PhD student in Economics Department at Fordham University. Email: wye22@fordham.edu

```

summary(olslwage)
stargazer(iv1, olslwage)
stargazer(olseduc)

#6.1 part b
u_hat1 <- resid(iv1)
# Then we regress estimated u1 on all the variables with educ replacing by
lm61b = lm(u_hat1 ~ exper + expersq + black + south + smsa + reg661 + reg662)

# Obtain R square value
Rsqr = summary(lm61b)$r.squared
pchisq(Rsqr * nrow(card), df=1, lower.tail=FALSE)

```

2 Question 2 – 6.2

Solution:

From the test, we can hold strong opinion that there exists exogeneity between these two variables.

```

#6.3
# Read the data
NLS80 = read_csv("NLS80.csv")
# Run the reduced forms
lm_rd1 = lm(educ ~ exper + tenure + married + south + urban + black + kww + medv)
v21_hat = resid(lm_rd1)
length(NLS80$educ)
length(v21_hat)
# Since the lengths of the response variable and the predictor are not the same
NLS80omit = na.omit(NLS80)
# Run the regression again
lm_rd1 = lm(educ ~ exper + tenure + married + south + urban + black + kww + medv)
v21_hat = resid(lm_rd1)

lm_rd2 = lm(iq ~ exper + tenure + married + south + urban + black + kww + medv)
v22_hat = resid(lm_rd2)

# Run the 2nd stage regression
lm_2s = lm(lwage ~ exper + tenure + married + south + urban + black + educ + v21_hat + v22_hat)
summary(lm_2s)

linearHypothesis(lm_2s, c("v21_hat=0", "v22_hat=0"))

```

3 Question 3 – 6.3

Solution:

- a) Price is correlated with calories and protein, but uncorrelated with u_1 .
- b) two unknown variable calories, and protein, thus, we at least need 2 equations.
- c) We would first estimate the two reduced forms for calories and protein by regressing each on a constant, $\text{exper}, \text{exper}^2$ and educ and the M prices p_1, \dots, p_M , we obtain $\hat{v}_{21}, \hat{v}_{22}$. Then we would run the regression $\log(\text{produc})$ on $1, \text{exper}, \text{exper}^2$ and educ, $\hat{v}_{21}, \hat{v}_{22}$, and do a joint significance test on $\hat{v}_{21}, \hat{v}_{22}$. We could use a standard F test or use a heteroskedasticity-robust test.

4 Question – 6.8

Solution:

- a) The estimate says that a women with about eight more years of education has about one fewer child, while keeping other factors fixed. The coefficient is also very statistically significant. There has been a notable secular decline in fertility over this period: on average, with other factors held fixed, a women in 1984 had about half a child less -0.545 than a similar woman in 1972, the base year. The effect is also statistically significant with p-value = .002.
- b) Yes, they are partially correlated with edu.
- c) t-value on \hat{v}_2 is .702, which means there is nearly no evidence that educ is endogenous. “ The estimated coefficient on educ is larger in magnitude than before, but the test for endogeneity shows that we can reasonably attribute the difference between OLS and 2SLS to sampling error.
- d)
 - Since there is little evidence that educ is endogenous, we could just use OLS. I did it both ways. First, I just added interactions $y74\text{educ}, y76\text{educ}, \dots, y84\text{educ}$ to the model in part a and used OLS.
 - Some of the interactions, particularly in the last two years, are marginally significant and negative, showing that the effect of education has become stronger over time.
 - But the joint F test for the interaction terms yields p-value = .180, and so we do not reject the model without the interactions. Still, the possibility that the link between fertility and education has become stronger over time is deserves attention, especially using more recent data.
 - Qualitatively, the results are similar to the OLS estimates. The p-value for the joint F test on the interactions is 0.204, which has asymptotic justification under Assumption 2SLS.3, the homoskedasticity assumption - so again there is

no strong evidence favoring including of the interactions of year dummies and education.

```
###partc
```

```
lm_68c1 = lm(kids ~ educ+age+agesq+black+east+northcen+west+farm+othrural+  
summary(lm_68c1)
```

```
linearHypothesis(lm_68c1, c("y74educ=0", "y76educ=0", "y78educ=0", "y80educ=0"))
```

```
lm_68c2= ivreg(kids ~ age+agesq+black+east+northcen+west+farm+othrural+town+  
summary(lm_68c2)
```

```
linearHypothesis(lm_68c2, c("y74educ=0", "y76educ=0", "y78educ=0", "y80educ=0"))
```

Appendix

Table 1: 6.1 Reduced Form of Educ

	<i>Dependent variable:</i>
	educ
exper	-0.410*** (0.034)
expersq	0.001 (0.002)
black	-1.013*** (0.090)
south	-0.279*** (0.080)
smsa	0.389*** (0.086)
reg661	-0.006 (0.171)
nearc2	0.108 (0.073)
nearc4	0.331*** (0.083)
Constant	16.623*** (0.178)
Observations	3,010
R ²	0.475
Adjusted R ²	0.473
Residual Std. Error	1.942 (df = 3001)
F Statistic	339.192*** (df = 8; 3001)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 2: 6.1: OLS w/ educhat and 2SLS comparison

	<i>Dependent variable:</i>	
	lwage	
	<i>instrumental variable</i>	<i>OLS</i>
	(1)	(2)
exper	0.122*** (0.021)	0.122*** (0.021)
expersq	-0.002*** (0.0004)	-0.002*** (0.0003)
black	-0.100* (0.053)	-0.100* (0.051)
south	-0.100*** (0.024)	-0.100*** (0.023)
smsa	0.112*** (0.031)	0.112*** (0.030)
reg661	-0.107*** (0.036)	-0.107*** (0.035)
educ	0.166*** (0.049)	
educhat		0.166*** (0.047)
Constant	3.190*** (0.827)	3.190*** (0.798)
Observations	3,010	3,010
R ²	0.130	0.191
Adjusted R ²	0.128	0.189
Residual Std. Error (df = 3002)	0.414	0.400
F Statistic		101.156*** (df = 7; 3002)
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01

Table 3: 6.8 Regression Results between 2SLS and OLS

	<i>Dependent variable:</i>	
	kids	educ
	(1)	(2)
educ	-0.128*** (0.018)	
age	0.532*** (0.138)	-0.224 (0.200)
agesq	-0.006*** (0.002)	0.003 (0.002)
black	1.076*** (0.174)	0.367 (0.252)
east	0.217 (0.133)	0.249 (0.192)
northcen	0.363*** (0.121)	0.091 (0.176)
west	0.198 (0.167)	0.101 (0.242)
farm	-0.053 (0.147)	-0.379* (0.214)
othrural	-0.163 (0.175)	-0.561** (0.255)
town	0.084 (0.125)	0.062 (0.181)
smcity	0.212 (0.160)	0.081 (0.232)
y74	0.268 (0.173)	0.006 (0.250)
y76	-0.097 (0.179)	0.124 (0.259)
y78	-0.069 (0.182)	0.208 (0.263)
y80	-0.071 (0.183)	0.383 (0.264)
y82	-0.522*** (0.172)	0.582** (0.249)