Homework 5

Wei Ye*
ECON 7910- Econometrics I

Due on Oct 21, 2021

1 Question – 5.1

Solution:

Like in the hint, we first make regression of y_2, π_2 on v_2, then save the residuals, we say $\ddot{x}$. Put the residuals into the equation of y_1, then still make linear regression to get the efficients of δ and y_2. The thought behind this is v_2 is uncorrected with y_2 and π_2.

Therefore, the OLS and 2SLS can have the same coefficient estimation values.

2 Question – 5.3

Solution:

a) Becasue other variables can also affect women's health level during pregancy, like exercise frequency, drugs etc.

b) i) Cigarette price is uncorrelated with the errors in (5.54), because we guess there may be other variables that correlated with packs, but it seems that ciga price doesn't affect exercise frequence and drugs.

ii) Cigarete price is obviously correlated with packs, in general, higher the prices of the cigareets are, less packs needed.

c) From Regression table (1), after adding an IV, the effect of packs doesn't only become from negavity to positivity, but also from significance to insignificance. Other variables also become insignificant, which means cigprice may not be a good IV.

d) From the table (2), we can't reject the coefficient of cigprice is equal to 0, so still cigprice is not a good IV for packs as the same with part (c).

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

```
        ######################
###These are the codes for HW-5
### All the codes are written by Wei Ye
## Date: Oct 20, 2021
install.packages("ivreg")
library(tidyverse)
library(stargazer)
library(ivreg)

#For Question 5.3
bwght<-read_csv('bwght.csv')
head(bwght,n=5)
bwght_ols<- lm(bwghtlbs~male+parity+lfaminc+packs,data=bwght)
summary(bwght_ols)
bwght_2sls<- ivreg(bwghtlbs~male+parity+lfaminc+packs | male+parity+lfaminc+c
summary(bwght_2sls)
stargazer(bwght_ols,bwght_2sls)
### 5.3 part(d) Reduced form
pack_ols <- lm(packs~male+parity+lfaminc+cigprice,data=bwght)
summary(pack_ols)
stargazer(pack_ols)
```

3 Question – 5.7

Solution:

a) We first assume $\xi = \frac{1}{\delta_1}$ From the information of this question, we can rewrite the equation to

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \xi q_1 - \xi a + v$$

We need to ask q_1 is uncorrelated with a_1, $x_1, ..., x_k, z_1, ..., z_M$ and v, and $\xi \neq 0$.

b) First, given IQ information, $z_1, ..., z_M$ can be overlooked, because, IQ can make other family background information redundant. Second, IQ should be correlated with family background in some sense, because although family background can't affect log wage direcly, it can affect q_1, like more educated parents tend to have more intelligent and well-educated children, and these children may perform well in tests.

c) It seems that IV is pretty good. Because educ is still strongly significant and other variables behave well on the at 95% significance level.

```
    ####################
###Question 5.7#####
####################
```

```
nls80 <- read_csv('nls80.csv')
head(nls80,head=5)
nls2sls1 <- ivreg(lwage~exper+tenure+educ+married+south+urban+black | exper+
                    tenure+married+south+urban+black+iq+meduc+feduc+sibs,
                  data=nls80)
summary(nls2sls1)
stargazer(nls2sls1)
```

4 Question – 5.9

Solution:

Rewrite H_0 to $\beta_4 - \beta_3 = 0$, and $H_1 : \beta_4 - \beta_3 > 0$. We can denote $\delta = \beta_4 - \beta_3$ for our convenience. $H_0 : \delta = 0$, and $H_1 : \delta > 0$. Then rewrite the wage equation:

$$\begin{aligned}
\log(wage) &= \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 twoyr + \beta_4 fouryr + u \\
&= \beta_0 + \beta_1 exper + \beta_2 exper^2 + (\beta_4 - \beta_3) fouryr + \beta_3 twoyr + \beta_3 fouryr + u \\
&= \beta_0 + \beta_1 exper + \beta_2 exper^2 + \delta foury + \beta_3 (twoyr + fouryr) + u \qquad (1)
\end{aligned}$$

Since two-year college and four-year college in general exclusively, we can form a new column in the data called $totalyr = twoyr + fouryr$), then make a regression on the above equation, to estimate the coefficient δ, if $\delta = 0$ is significant, we may say H_0 is correct, if $\delta > 0$ in 95% confidence level, we say, H_1 is accepted.

Appendix

Table 1: 5.3 IV and 2SLS Regression Results

	Dependent variable:	
	bwghtlbs	
	OLS	*instrumental variable*
	(1)	(2)
male	0.197***	0.223*
	(0.067)	(0.125)
parity	0.104***	−0.011
	(0.038)	(0.154)
lfaminc	0.129***	0.458
	(0.037)	(0.401)
packs	−0.599***	5.764
	(0.114)	(7.640)
Constant	6.813***	5.313***
	(0.146)	(1.820)
Observations	1,388	1,388
R^2	0.041	−2.124
Adjusted R^2	0.038	−2.133
Residual Std. Error (df = 1383)	1.248	2.252
F Statistic	14.631*** (df = 4; 1383)	
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 2: 5.3 Reduced Form Regression

	Dependent variable:
	packs
male	−0.005
	(0.016)
parity	0.018**
	(0.009)
lfaminc	−0.053***
	(0.009)
cigprice	0.001
	(0.001)
Constant	0.137
	(0.104)
Observations	1,388
R^2	0.030
Adjusted R^2	0.028
Residual Std. Error	0.294 (df = 1383)
F Statistic	10.860*** (df = 4; 1383)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3: 5.7 Part c Regression Results

	Dependent variable:
	lwage
exper	0.027***
	(0.005)
tenure	0.007**
	(0.003)
educ	0.112***
	(0.013)
married	0.204***
	(0.045)
south	−0.074**
	(0.031)
urban	0.173***
	(0.032)
black	−0.119**
	(0.052)
Constant	4.646***
	(0.222)
Observations	722
R^2	0.195
Adjusted R^2	0.187
Residual Std. Error	0.378 (df = 714)
Note:	*p<0.1; **p<0.05; ***p<0.01