Obaid Masih

Econometrics HW

Assignment 8 - Instruments Variables

Question 10.2 (p.424)

HOURS = b1 + b2 WAGE + b3 EDUC + b4 AGE + b5 KIDSL6 + b6 KIDS618 + b7 NWIFEINC + e

a)

B2 is expected to be positive. As Wage increases then Hours are also expected to increase. But this is not expected to be a linear relationship.

B3 Educ is expected to be positive. As more education can get you more responsibility. Relation is not expected to be linear but polynomial.

B4 Age is expected to be negative. Meaning more Age can limit your working hours.

B5 is expected to be negative. More kids will mean less hours at work and more hours spent taking care of children under 6

B6 is also expected to be negative. In my own experience children between 6-18 do require a lot of attention and that will result in less hours of work at the job

B7 is also expected to be negative. If income from other source is large then wife will not have a dire need to work too many hours.

b) Equation cannot consistently estimate because one of the x variables is WAGE. Which is endogenous because it is simultaneously bias to labor hours. It is highly correlated to error which can yield bias and inconsistent estimates.

C) EXP and EXP^2 will be good instrument variables because of the following reasons:

1. They do not have a direct effect on *HOURS*, and thus it does not belong on the right-hand side of the model as an explanatory variable
2. They are not correlated with the regression error term *e*
3. *They are* strongly [or at least not weakly] correlated with the endogenous explanatory variable WAGE

D) Supply equation is identified as we have two IV against one endogenous variable. Estimates will be over identified as IV > EV. But this still meets the criteria as we should have Instrumental variables should be more than or equal to Endogenous variables in the system.

E) In first step we will introduce the first endogenous variable on the LHS and all variables including two instrument variables (EXP & EXP^2) on the RHS. Once the model is complete then we will run a hypothesis test (strength test) on the estimate of the two instrument variables.

HO: ALPHA 1 =0 , alpha 2 = 0 . Where alpha1 & 2 are the estimates for EXP and EXP^2

H1 : At least one of them is not 0

If F TEST is more than 10 only then we can reject the HO and accept EXP and EXP^2 as instrument variables.

In second step we will run our model again for hours. Only this time we will replace our two endogenous variables with two Exogenous variables.

Question 10.6 (p.425-426)

(part a)

$$\text{var}\left(\hat{\beta}_2\right) = \frac{\sigma^2}{r_{zx}^2 \sum \left(x_i - \bar{x}\right)^2} = \frac{\text{var}\left(b_2\right)}{r_{zx}^2}$$

X mean = 0 and variance = 2

Error e mean =0 and variance = 1

Cov(x,e) = 0.9

$$Correlation = \rho = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Corr (x, e) = cov(x,e) / sqrt (var(x) * var(e))

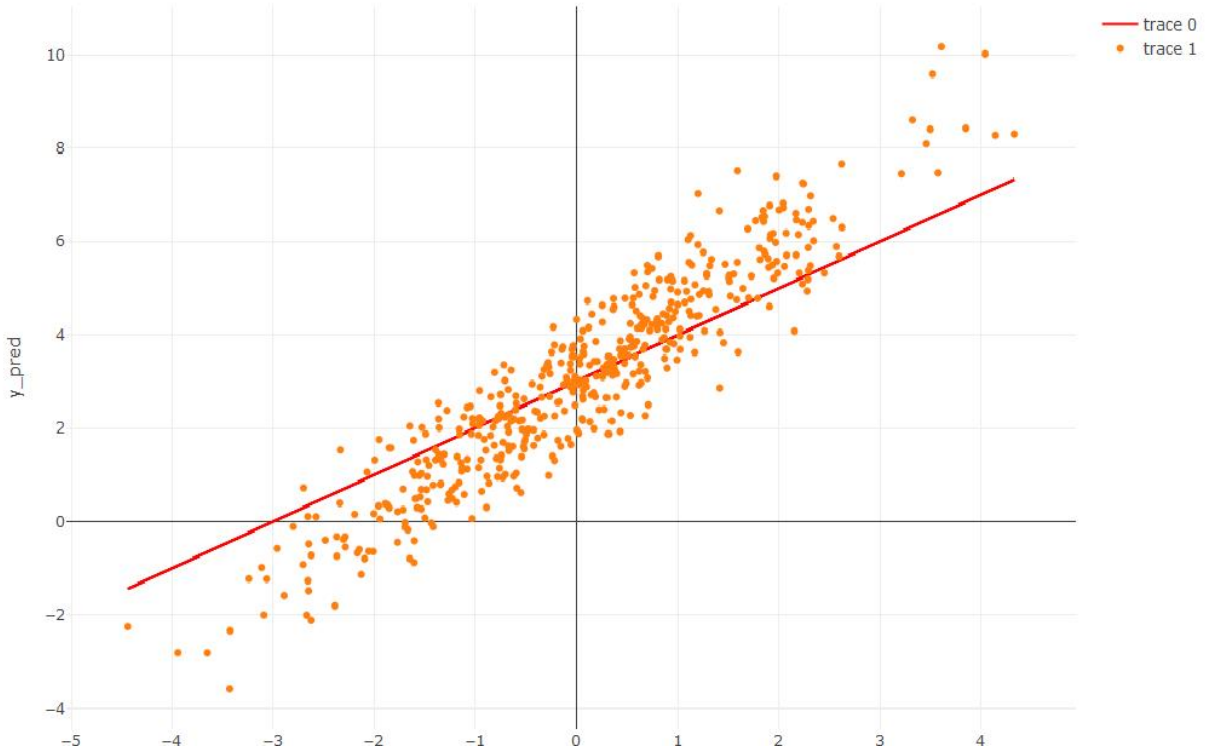
= 0.9 / sqrt(2 *1) = 0.63

(part b)

Correlation from R studios.

cor(context$x,context$e) = 0.65

It is very close to the value calculate above



Above plot is made in R Studio

Orange scatter plot is for the original y vs x

Red line is the y=3 +x

Plot shows that when X is negative then points fall below the regression line.

When X is positive then points are above the regression line.

(part c)

Running regression y= B1 + B2x

N=10

Intercept = 2.7775

x = 1.3722

N=20

Intercept = 3.0169

x = 1.387

N = 100

Intercept = 3.007

x = 1.40

N = 500

Intercept = 3.018

X = 1.45

It is evident that increasing sample size moves X estimate away from the true value. But Intercept goes closer to the true value. This is telling us that intercept estimate is consistent but X estimate are biased and inconsistent.

(part d)

Correlation found in R studio

cor(context$x,context$z1)

Correlation between X & Z1 0.6208

cor(context$z1,context$e)

Correlation found between Z1 & error -0.003447192

cor(context$x,context$z2)

Correlation found between X and Z2 0.289

cor(context$z2,context$e)

Correlation found between Z2 and error 0.02777

This reveals that Z1 is a better variable as it has a high correlation with X and no correlation with error of the regression.

(part f)

In this part we convert X to exogenous by using Z1 as the IV . This was done in R studio with library AER and function Ivreg()

Results are as follows

Sample Size	Intercept	X value
N = 10	2.7144	1.0640
N = 20	3.0810	1.0263
N = 100	2.9771	0.9363
N =500	3.03	0.99613

This shows that as sample size increases then Intercept and X value both move closer to the true value. This means our model is consistent

(part g)

In this part we convert X to exogenous by using Z2 as the IV . This was done in R studio with library AER and function Ivreg()

Results are as follows

Sample Size	Intercept	X value
N = 10	1.8913	-2.95
N = 20	3.24	0.11
N = 100	2.990	1.13
N =500	3.021	1.06

Introducing Z2 as the IV to the model has made X consistent. As X values increase then X estimate converges with the true value.

However estimates are wrong when sample size is small. But increasing sample size resolves the issue. IV Z1 predicted X estimate and Intercept better than Z2

(part h)

In this part we convert X to exogenous by using Z1 & Z2 as the IV . This was done in R studio with library AER and function Ivreg()

Results are as follows

Sample Size	Intercept	X value
N = 10	2.711	1.0491
N = 20	3.085	1.002
N = 100	2.980	0.9920
N =500	3.03	1.008

Results are consistent because increasing the sample size make intercept and X estimate converge to true value. Even at low sample size the predictions are consistent and close to true values.

Results are slightly better when Z1 was used as IV.

********** CODE FOR ABOVE QUESTION/S **********

```
##############################################
## QUESTION 8.12
rm(list=ls(all=TRUE))



#library(multcomp)
library(data.table)
library(dplyr)
library(plotly)
#library(lmtest)
#library(sandwich)
#library(car)
library(AER)

context <- fread("ivreg2.csv")

# part b   use b1 = 3 and b2 = 1  find e

context<- mutate(context, y_pred= 3+x )
context <- mutate(context, e= y - y_pred)

cor(context$x,context$e)

## part c  plot the graphs
```

```
plot_ly(context)%>%
add_trace(x = ~x, y = ~y_pred, type = 'scatter', mode = 'lines', line = list(color = 'red')) %>%
add_trace(x=~x, y = ~y)

## part d

# N=10
context10 <- context[0:10,]
model10 <- lm(y~x,context10)
summary(model10)

#Coefficients:
#Estimate Std. Error t value Pr(>|t|)
#(Intercept)   2.7775     0.3608   7.698 5.76e-05 ***
#  x           1.3722     0.1727   7.945 4.59e-05 ***

# N=20

context20 <- context[0:20,]
model20 <- lm(y~x,context20)
summary(model20)

#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept)   3.0169     0.2036   14.81 1.59e-11 ***
#  x           1.3876     0.1211   11.46 1.05e-09 ***

# N=100
```

```
context100 <- context[0:100,]
model100 <- lm(y~x,context100)
summary(model100)

#Coefficients:
 # Estimate Std. Error t value Pr(>|t|)
#(Intercept)  3.00783    0.07872   38.21   <2e-16 ***
#  x          1.40164    0.05330   26.30   <2e-16 ***

# N=500

context500 <- context[0:500,]
model500 <- lm(y~x,context500)
summary(model500)

#Coefficients:
#  Estimate Std. Error t value Pr(>|t|)
#(Intercept)  3.01825    0.03410    88.5   <2e-16 ***
#  x          1.45352    0.02367    61.4   <2e-16 ***

# part e

cor(context$x,context$z1)
# 0.6208
cor(context$z1,context$e)
# -0.003447192

cor(context$x,context$z2)
```

```
# 0.289
cor(context$z2,context$e)
#0.02777

## Z1 is  a better choice

## Part f lets use z1 as the instrumental variable
##################################################
##################################################
# USE Z1 as instrumental variable



# n=10
modelz1 <- ivreg(y~x|z1,data=context[0:10,])
summary(modelz1)

# n=20
modelz1 <- ivreg(y~x|z1,data=context[0:20,])
summary(modelz1)

# n=100

modelz1 <- ivreg(y~x|z1,data=context[0:100,])
summary(modelz1)

#n=500
modelz1 <- ivreg(y~x|z1,data=context[0:500,])
summary(modelz1)
```

```
##################
#part f
###  use z2 as the instrument variable

# n=10
modelz2 <- ivreg(y~x|z2, data = context[0:10,])
summary(modelz2)

# n=20
modelz2 <- ivreg(y~x|z2, data = context[0:20,])
summary(modelz2)

# n=100
modelz2 <- ivreg(y~x|z2, data = context[0:100,])
summary(modelz2)

# n=500
modelz2 <- ivreg(y~x|z2, data = context[0:500,])
summary(modelz2)

###############################
# part h
### using z1 and z2 as the instrument variable

# n=10
modelz1.z2 <- ivreg(y~x|z1+z2, data = context[0:10,])
summary(modelz1.z2)

# n=20
```

```
modelz1.z2 <- ivreg(y~x|z1+z2, data = context[0:20,])
summary(modelz1.z2)

# n=100
modelz1.z2 <- ivreg(y~x|z1+z2, data = context[0:100,])
summary(modelz1.z2)

# n=500
modelz1.z2 <- ivreg(y~x|z1+z2, data = context[0:500,])
summary(modelz1.z2)
```