cls

//Answers to the Computer Exercises in week 2: Natural experiments: Instrumental Variables

clear all

//open data folder (change to YOURFOLDER, so type in the location of your folder between the quotation marks)

cap cd "C:\Users\user\Desktop\master\Semester 1\Applied financial econometrics\Tutorial\w2"

/\*

// \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

//PART A

//1

set seed 4444

set obs 100

//2

generate u = invnorm(uniform())

generate i = invnorm(uniform())

generate c = 10+i+2\*u

generate y = 10+2\*i+2\*u

//3

regress c y, robust

// we see that the estimated effect of Y is .6922221 and that 0.5 is far left of

// the 95% CI; We know the true value is 0.5 which is the first indication that the OLS estimator is biased

// which you can verify by doing simulations (see q 5)

//4

test y==0.5

// the test that OLS gives us the true value of 0.5 is clearly rejected and will probably always be rejected

// 5 and 6: For the repetitions, we will use the syntax forv i=1/8 { commands }

// that tells stata to repeat the commands 8 times. The first four repetitions will be with 100 obs,

// the last four with 1000 observations, because that is what part 6 asks you to do

forv i = 1 / 8 {

qui drop \_all

qui set obs 200

if (`i' > 4 ) qui set obs 1000 // this is to set the obs to 1000 for part 6

di \_n \_n \_n as text "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*Round `i'" // command to show simulation round on screen

generate u = invnorm(uniform())

generate i = invnorm(uniform())

generate c = 10+i+2\*u

generate y = 10+2\*i+2\*u

regress c y, robust

test y==0.5

}

//Part 5 remarks:

// The results of Round 1, 2, 3 and 4 show that OLS is always far off at aprox 0.75, with rejection of the true value

//Part 6 remarks:

// The results of Round 5, 6, 7 and 8 show that even with a large sample,

// OLS is far off at aprox 0.75

//7

ivregress 2sls c (y = i), robust

// the point estimate is now very close to 0.5, and this value lies nicely

// the 95% confidence interval. TSLS worked!

//8

test y=0.5

//the test that IV gives us the true value of 0.5 is not rejected (which doesnt mean we accept it,

// but if you were to repeat this test you would only reject in 5% of cases, i.e. the appropriate

// number of times; you find this out by simulating data as we did in the previous exercise)

//9

//Manual IV; don't try this at home, you get wrong s.e.'s!

regress y i, robust

predict yfit

regress c yfit, robust

// we see that you get the same estimate for beta1 as in q7, but the estimated s.e. is

// not equal to the s.e. of TSLS, which is the correct one (i.e. consistent).

//10

clear

\*/

//\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

//PART B

//11; open data

use w2\_brabant.dta, clear

//12

reg lwage educ c.lexp##c.lexp, robust

est sto ols

// Semi elasticity: a one unit change in educ (i.e. one additional year of education)

// leads to a 5% change in wages (because wage is in logarithms)

//13

ivregress 2sls lwage c.lexp##c.lexp (educ = faed mark ssoc fhigh fint fself), robust

est sto iv

// Dummy variable trap: we cannot include FLOW because FINT, FHIGH and a constant term are all included.

// Exogeneity of the instrument(s) is always open to discussion. In general fathers education and social background will probably

// not only affect schooling but also affect your wages in other ways (i.e. the skills learned at home) and therefore will

// probably also have effects of themselves (direct effect: low ssoc -> low social skill & network -> worse labor opportunities and outcomes, not only

// due to lower schooling) or be correlated to other determinants of wages (father self-employed will probably relate to living location for instance

// (i.e favourable local conditions for entrepreneurs), which also affects the job and wages of their children); Therefore these instruments

// are not very credible.

//14

regress educ faed mark ssoc fhigh fint fself c.lexp##c.lexp, robust

testparm educ faed mark ssoc fhigh fint fself

//F-stat=26.11 > 10 so we do not have a weak instrument problem

//15

est restore iv

estat firststage

//again F-stat=26.11 > 10 so we do not have a weak instument problem, now given by Stata directly; The partial

// R2= 0.19, which is the variance explained by all the instuments together.

//16

//cap ssc install estout

esttab ols iv, b(%6.4f) se(%6.4f)

// IV estimates are higher, move from 5%->9%

// this is surprising given that we would expect that OLS is upward biased due to positive selection;

// with "smart" individuals obtaining educ having better wage perspectives in any case.

// There may be several reasons that IV>OLS; in this case it is likely that the instrument fails because

// fathers will pass on part of their IQ to their kids. Hence instrument exogeneity probably fails.

estat endogenous

// with an F(1,834) = 8.42304 and pvalue = 0.0038, the OLS is clearly rejected; This means we have evidence that OLS

// is biased (downward)

//17

estat overid

// reject H0 that instruments do not contradict each other because J=mF=10.08 > Chi2-critical=9.24

// we accept the alternative hypothesis that the instuments contradict each other, but it is unclear

// what this tells us exactly.