clear

clear all

set more off, perm

cd "E:\MSc\2018-2019\Labs\24-11-2018"

log using lab2.txt, text replace

\*\*\*SECTION A: Artificial Data

\*\*\*Sample size and weak IVs

\*\*Sample size (observations: 50, 500, 1000, 2000, 5000)

local N = 5000

local seedn = 457387+`N'

set seed `seedn'

set obs `N'

\*\*Define instrument

gen z = invnorm(uniform())

kdensity z

kdensity z, normal

scalar std\_e1 = 1

scalar std\_v2 = 1

gen e1 = std\_e1\*invnorm(uniform())

gen e2 = std\_v2\*invnorm(uniform())

kdensity e1

kdensity e2

gen u1 = e1 + e2

kdensity u1

gen x = 1\*z + e2

kdensity x

gen y = 0.5\*x + u1 /\*beta: 0, 0.5; error: e1, u1\*/

kdensity y

reg x z

est store m1FirstS

reg y x

est store m1RedForm

ivreg2 y (x = z), first

est store m1IV2SLS

estimates table m1FirstS m1RedForm m1IV2SLS, keep(x z) b(%9.3g) se t

\*\*\*SECTION B: REAL DATA (Card)

/\*

Card (1995), “Using geographic variation in college proximity to estimate the return

to schooling”, published in L. Christophides, E. Grant and R. Swidinsky (eds.),

Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp

\*/

clear

clear all

set more off, perm

use "http://fmwww.bc.edu/ec-p/data/wooldridge/card.dta", clear

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*One instrument (Table 1 to Table 4)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*Table 1 STATS

tabstat lwage educ nearc2 nearc4, s(mean N min max p50)

\*\*Table 2 OLS

reg lwage educ

\*\*Table 3 IV-2SLS

ivreg2 lwage (educ = nearc4), first // SEE COMMENT AT THE END OF THE DO FILE

\*\*Table 4 Wald Estimate

tabstat lwage educ, s(mean) by(nearc4)

display (6.311401- 6.155494)/(13.52703-12.69801)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*Two instruments (Table 1 to Table 3, Exactly the same as Table 4)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*Table 1 OLS

reg lwage educ exper expersq black south smsa ///

reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668 smsa66

\*\*Table 2 Reduced form education for education

reg educ nearc2 nearc4 motheduc fatheduc ///

exper expersq black south smsa ///

reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668 smsa66

\*\*F-test

test nearc2 nearc4 motheduc fatheduc

\*\*Get residuals

predict residuals, res

\*\*Table 3 Regression based Hausman test for endogeneity

reg lwage educ exper expersq black south smsa ///

reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668 smsa66 residuals

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*Table 4 2SLS estimates

ivreg2 lwage (educ = nearc2 nearc4 motheduc fatheduc) exper expersq black ///

south smsa reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668 smsa66

ivendog //Provides an endogeneity test exactly like the -endog- option

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*Table 5: 2SLS estimates excluding parents' education

ivreg2 lwage (educ = nearc2 nearc4) exper expersq black south smsa reg661 ///

reg662 reg663 reg664 reg665 reg666 reg667 reg668 smsa66

ivendog

//The 'first' option of ivreg2 presents several useful

//diagnostics that assess the first-stage regressions. If there is a single

//endogenous regressor, these issues are simplified, as the instruments

//either explain a reasonable fraction of that regressor’s variability or not.

//With multiple endogenous regressors, diagnostics are more

//complicated, as each instrument is being called upon to play a role in

//each first-stage regression. With sufficiently weak instruments, the

//asymptotic identification status of the equation is called into question.

//The 'Sargan statistic' is an overidentification test of all instruments.

//Or a simple overidentification test

//It can also be calculated after ivreg2 estimation with the

//overid command, which is part of the ivreg2 suite

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*EXAMPLE\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

ivreg2 lwage (educ = nearc2 nearc4) exper expersq black south smsa reg661 ///

reg662 reg663 reg664 reg665 reg666 reg667 reg668 smsa66

ssc install overid //it is a user-written command, therefore it should be installed

overid

log close