. \*-----------------------------------\*

. \* Variables instrumentales \*

. \*-----------------------------------\*

.

. \*-----------------\*

\*En el capítulo de variables instrumentales se utilizará el comando "ivreg2". Dado que este comando no viene

\*incluido en el paquete básico de Stata, es necesario instalarlo. Para instalarlo debemosseleccionar

. \*la pestaña "help" de Stata, en la opción search y seleccionamos

. \*search all. Buscamos "ivreg2" y seleccionamos el paquete st0030\_2 para instalarlo.

. \*-----------------------\*

.

Para simplificar el análisis definimos un vector global con las variables que se han utilizado antes:

.

. global X "personas orden\_n ocupado\_jefe educa\_jefe ingresos\_hogar\_jefe"

.

Cuando un regresor tiene alguna relación con el término de error, se viola uno de los supuestos ásicos

de MCO y por lo tanto, estimaciones mediante este método arrojarían resultados sesgados. En nuestro ejemplo, el tratamiento puede estar relacionado con variables no observables como el nivel de preocupación de una madre frente a su hijo o el conocimiento que ésta tiene sobre el desarrollo adecuado de los niños

. \*\*\*\* Veamos los resultados de una estimación simple por MCO:

.

. reg ha\_nchs D $X

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 6, 3993) = 39.49

Model | 68.9063178 6 11.4843863 Prob > F = 0.0000

Residual | 1161.23317 3993 .290817222 R-squared = 0.0560

-------------+------------------------------ Adj R-squared = 0.0546

Total | 1230.13949 3999 .307611775 Root MSE = .53927

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2489481 .0171935 14.48 0.000 .2152392 .282657

personas | -.0093891 .004544 -2.07 0.039 -.018298 -.0004802

orden\_n | -.0632222 .0263134 -2.40 0.016 -.1148111 -.0116333

ocupado\_jefe | -.0364222 .0223229 -1.63 0.103 -.0801875 .0073431

educa\_jefe | .005974 .002325 2.57 0.010 .0014157 .0105322

ingresos\_h~e | .0001628 .000101 1.61 0.107 -.0000353 .0003609

\_cons | -.0840199 .0452452 -1.86 0.063 -.1727257 .004686

------------------------------------------------------------------------------

.

. \*\*\*\* Los resultados nos indican que el tratamiento tiene un efecto de 0.248 en la talla para la edad

. \*\*\*\* de los individuos. Sin embargo, dado el problema de autoselección, este resultado puede estar ses

> gado.

. \*\*\*\* Para proceder a utilizar la metodología de variables instrumentales debemos seleccionar instrumen

> tos

. \*\*\*\* que sean relevantes y exógenos. En principio, el número de oficinas operadoras en el municipio de

> termina

. \*\*\*\* en cierta medida la probabilidad de participación sin tener relación con las variables no observa

> bles.

. \*\*\*\* Lo mismo sucede con la distancia del hogar a la oficina operadora más cercana. Antes de utilizar

. \*\*\*\* estos instrumentos realizamos pruebas de relevancia y exogeneidad:

.

. \*------------------\*

. \* 1. Relevancia \*

. \*------------------\*

.

. \*\*\*\* Para evaluar la relevancia de un instrumento se pueden utilizar diferentes pruebas. En este caso

> vamos a utilizar

. \*\*\*\* una estimación por MCO, la prueba canónica de Anderson, la prueba de Cragg-Donald y la prueba de

> Stock y Yoko.

.

. \*\*\*\* 1.1 Distancia

.

. \*\*\*\* 1.1.1 Regresión MCO

.

. reg D distancia

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 1, 3998) = 50.14

Model | 12.3778733 1 12.3778733 Prob > F = 0.0000

Residual | 987.046127 3998 .246884974 R-squared = 0.0124

-------------+------------------------------ Adj R-squared = 0.0121

Total | 999.424 3999 .24991848 Root MSE = .49688

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

distancia | -.0000498 7.03e-06 -7.08 0.000 -.0000636 -.000036

\_cons | .5889147 .016274 36.19 0.000 .5570085 .6208208

------------------------------------------------------------------------------

.

. reg D $X distancia

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 6, 3993) = 19.23

Model | 28.0671549 6 4.67785916 Prob > F = 0.0000

Residual | 971.356845 3993 .243264925 R-squared = 0.0281

-------------+------------------------------ Adj R-squared = 0.0266

Total | 999.424 3999 .24991848 Root MSE = .49322

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.0211905 .0041425 -5.12 0.000 -.0293121 -.013069

orden\_n | .0937512 .0240311 3.90 0.000 .0466369 .1408655

ocupado\_jefe | .0785951 .020379 3.86 0.000 .0386409 .1185493

educa\_jefe | -.0058249 .0021244 -2.74 0.006 -.0099899 -.0016598

ingresos\_h~e | .0003046 .0000923 3.30 0.001 .0001237 .0004856

distancia | -.0000499 6.98e-06 -7.14 0.000 -.0000635 -.0000362

\_cons | .5447223 .0429713 12.68 0.000 .4604746 .62897

------------------------------------------------------------------------------

.

. \*\*\*\* Vemos que la distancia sí puede ser un buen predictor de la variable "D".

. \*\*\*\* El valor del estadístico F es 19,23. Según Stock y Watson pg. 443, (2.007), si el estadístico

. \*\*\*\* F es superior a diez, se puede asegurar que el instrumento es relevante. Además, todos los coefic

> ientes

. \*\*\*\* son estadísticamente significativos.

.

. \*\*\*\* 1.1.2 Prueba canónica de Anderson

.

. ivreg2 ha\_nchs $X (D=distancia)

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 4.64

Prob > F = 0.0001

Total (centered) SS = 1230.139487 Centered R2 = 0.0464

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0567

Residual SS = 1173.111503 Root MSE = .5416

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .1390646 .1537524 0.90 0.366 -.1622846 .4404138

personas | -.0117333 .0056077 -2.09 0.036 -.0227241 -.0007424

orden\_n | -.053356 .029773 -1.79 0.073 -.1117099 .004998

ocupado\_jefe | -.0275681 .0255749 -1.08 0.281 -.077694 .0225578

educa\_jefe | .0053271 .0025021 2.13 0.033 .0004231 .010231

ingresos\_h~e | .0001958 .0001113 1.76 0.079 -.0000224 .0004139

\_cons | -.0348009 .0821434 -0.42 0.672 -.1957989 .1261971

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 50.443

Chi-sq(1) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 50.998

Stock-Yogo weak ID test critical values: 10% maximal IV size 16.38

15% maximal IV size 8.96

20% maximal IV size 6.66

25% maximal IV size 5.53

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.000

(equation exactly identified)

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: distancia

------------------------------------------------------------------------------

.

. \*\*\*\* La hipótesis nula de esta prueba es que la ecuación está subidentificada. Dado que la rechazamos,

> la

. \*\*\*\* prueba canónica de Anderson (Underidentification test) nos permite estar más seguros de la releva

> ncia de nuestro instrumento.

.

. \*\*\*\* 1.1.3 Prueba Cragg-Donald

.

. ivreg2 ha\_nchs $X (D=distancia), first

First-stage regressions

-----------------------

First-stage regression of D:

OLS estimation

--------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 19.23

Prob > F = 0.0000

Total (centered) SS = 999.424 Centered R2 = 0.0281

Total (uncentered) SS = 1952 Uncentered R2 = 0.5024

Residual SS = 971.3568451 Root MSE = .4932

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.0211905 .0041425 -5.12 0.000 -.0293121 -.013069

orden\_n | .0937512 .0240311 3.90 0.000 .0466369 .1408655

ocupado\_jefe | .0785951 .020379 3.86 0.000 .0386409 .1185493

educa\_jefe | -.0058249 .0021244 -2.74 0.006 -.0099899 -.0016598

ingresos\_h~e | .0003046 .0000923 3.30 0.001 .0001237 .0004856

distancia | -.0000499 6.98e-06 -7.14 0.000 -.0000635 -.0000362

\_cons | .5447223 .0429713 12.68 0.000 .4604746 .62897

------------------------------------------------------------------------------

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe distancia

------------------------------------------------------------------------------

Partial R-squared of excluded instruments: 0.0126

Test of excluded instruments:

F( 1, 3993) = 51.00

Prob > F = 0.0000

Summary results for first-stage regressions

-------------------------------------------

Variable | Shea Partial R2 | Partial R2 | F( 1, 3993) P-value

D | 0.0126 | 0.0126 | 51.00 0.0000

Underidentification tests

Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)

Ha: matrix has rank=K1 (identified)

Anderson canon. corr. N\*CCEV LM statistic Chi-sq(1)=50.44 P-val=0.0000

Cragg-Donald N\*CDEV Wald statistic Chi-sq(1)=51.09 P-val=0.0000

Weak identification test

Ho: equation is weakly identified

Cragg-Donald Wald F-statistic 51.00

See main output for Cragg-Donald weak id test critical values

Weak-instrument-robust inference

Tests of joint significance of endogenous regressors B1 in main equation

Ho: B1=0 and overidentifying restrictions are valid

Anderson-Rubin Wald test F(1,3993)=0.78 P-val=0.3760

Anderson-Rubin Wald test Chi-sq(1)=0.79 P-val=0.3755

Stock-Wright LM S statistic Chi-sq(1)=0.79 P-val=0.3756

Number of observations N = 4000

Number of regressors K = 7

Number of instruments L = 7

Number of excluded instruments L1 = 1

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 4.64

Prob > F = 0.0001

Total (centered) SS = 1230.139487 Centered R2 = 0.0464

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0567

Residual SS = 1173.111503 Root MSE = .5416

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .1390646 .1537524 0.90 0.366 -.1622846 .4404138

personas | -.0117333 .0056077 -2.09 0.036 -.0227241 -.0007424

orden\_n | -.053356 .029773 -1.79 0.073 -.1117099 .004998

ocupado\_jefe | -.0275681 .0255749 -1.08 0.281 -.077694 .0225578

educa\_jefe | .0053271 .0025021 2.13 0.033 .0004231 .010231

ingresos\_h~e | .0001958 .0001113 1.76 0.079 -.0000224 .0004139

\_cons | -.0348009 .0821434 -0.42 0.672 -.1957989 .1261971

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 50.443

Chi-sq(1) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 50.998

Stock-Yogo weak ID test critical values: 10% maximal IV size 16.38

15% maximal IV size 8.96

20% maximal IV size 6.66

25% maximal IV size 5.53

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.000

(equation exactly identified)

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: distancia

------------------------------------------------------------------------------

.

. \*\*\*\* La prueba Cragg-Donald tiene la misma hipótesis nula que la prueba canónica de Anderson pero

. \*\*\*\* el estadístico de prueba se construye de manera diferente. En este caso, también rechazamos

. \*\*\*\* la hipótesis nula.

.

. \*\*\*\* 1.1.4 Prueba de Stock y Yoko

.

. ivreg2 ha\_nchs $X (D=distancia), first

First-stage regressions

-----------------------

First-stage regression of D:

OLS estimation

--------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 19.23

Prob > F = 0.0000

Total (centered) SS = 999.424 Centered R2 = 0.0281

Total (uncentered) SS = 1952 Uncentered R2 = 0.5024

Residual SS = 971.3568451 Root MSE = .4932

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.0211905 .0041425 -5.12 0.000 -.0293121 -.013069

orden\_n | .0937512 .0240311 3.90 0.000 .0466369 .1408655

ocupado\_jefe | .0785951 .020379 3.86 0.000 .0386409 .1185493

educa\_jefe | -.0058249 .0021244 -2.74 0.006 -.0099899 -.0016598

ingresos\_h~e | .0003046 .0000923 3.30 0.001 .0001237 .0004856

distancia | -.0000499 6.98e-06 -7.14 0.000 -.0000635 -.0000362

\_cons | .5447223 .0429713 12.68 0.000 .4604746 .62897

------------------------------------------------------------------------------

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe distancia

------------------------------------------------------------------------------

Partial R-squared of excluded instruments: 0.0126

Test of excluded instruments:

F( 1, 3993) = 51.00

Prob > F = 0.0000

Summary results for first-stage regressions

-------------------------------------------

Variable | Shea Partial R2 | Partial R2 | F( 1, 3993) P-value

D | 0.0126 | 0.0126 | 51.00 0.0000

Underidentification tests

Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)

Ha: matrix has rank=K1 (identified)

Anderson canon. corr. N\*CCEV LM statistic Chi-sq(1)=50.44 P-val=0.0000

Cragg-Donald N\*CDEV Wald statistic Chi-sq(1)=51.09 P-val=0.0000

Weak identification test

Ho: equation is weakly identified

Cragg-Donald Wald F-statistic 51.00

See main output for Cragg-Donald weak id test critical values

Weak-instrument-robust inference

Tests of joint significance of endogenous regressors B1 in main equation

Ho: B1=0 and overidentifying restrictions are valid

Anderson-Rubin Wald test F(1,3993)=0.78 P-val=0.3760

Anderson-Rubin Wald test Chi-sq(1)=0.79 P-val=0.3755

Stock-Wright LM S statistic Chi-sq(1)=0.79 P-val=0.3756

Number of observations N = 4000

Number of regressors K = 7

Number of instruments L = 7

Number of excluded instruments L1 = 1

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 4.64

Prob > F = 0.0001

Total (centered) SS = 1230.139487 Centered R2 = 0.0464

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0567

Residual SS = 1173.111503 Root MSE = .5416

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .1390646 .1537524 0.90 0.366 -.1622846 .4404138

personas | -.0117333 .0056077 -2.09 0.036 -.0227241 -.0007424

orden\_n | -.053356 .029773 -1.79 0.073 -.1117099 .004998

ocupado\_jefe | -.0275681 .0255749 -1.08 0.281 -.077694 .0225578

educa\_jefe | .0053271 .0025021 2.13 0.033 .0004231 .010231

ingresos\_h~e | .0001958 .0001113 1.76 0.079 -.0000224 .0004139

\_cons | -.0348009 .0821434 -0.42 0.672 -.1957989 .1261971

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 50.443

Chi-sq(1) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 50.998

Stock-Yogo weak ID test critical values: 10% maximal IV size 16.38

15% maximal IV size 8.96

20% maximal IV size 6.66

25% maximal IV size 5.53

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.000

(equation exactly identified)

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: distancia

------------------------------------------------------------------------------

.

. \*\*\*\* Para esta prueba utilizamos el estadístico Cragg-Donald y nos remitimos al valor reportado en la

> tabla de Stock y Yoko (2002).

.

. \*Realizamos las mismas pruebas para el instrumento "of\_op":

.

. \*\*\*\* 1.2 of\_op

.

. \*\*\*\* 1.2.1 Regresión MCO

.

. reg D of\_op

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 1, 3998) = 78.32

Model | 19.2033928 1 19.2033928 Prob > F = 0.0000

Residual | 980.220607 3998 .245177741 R-squared = 0.0192

-------------+------------------------------ Adj R-squared = 0.0190

Total | 999.424 3999 .24991848 Root MSE = .49515

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

of\_op | .0337162 .0038097 8.85 0.000 .0262471 .0411853

\_cons | .3983992 .0127982 31.13 0.000 .3733075 .4234909

------------------------------------------------------------------------------

.

. reg D $X of\_op

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 6, 3993) = 24.03

Model | 34.8331886 6 5.80553144 Prob > F = 0.0000

Residual | 964.590811 3993 .241570451 R-squared = 0.0349

-------------+------------------------------ Adj R-squared = 0.0334

Total | 999.424 3999 .24991848 Root MSE = .4915

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.021453 .004128 -5.20 0.000 -.0295461 -.0133598

orden\_n | .0934295 .0239443 3.90 0.000 .0464852 .1403738

ocupado\_jefe | .0772656 .0203094 3.80 0.000 .0374479 .1170834

educa\_jefe | -.0057974 .002117 -2.74 0.006 -.009948 -.0016469

ingresos\_h~e | .0003029 .000092 3.29 0.001 .0001226 .0004832

of\_op | .0336993 .0037827 8.91 0.000 .026283 .0411156

\_cons | .3567943 .0419026 8.51 0.000 .2746417 .4389468

------------------------------------------------------------------------------

.

. \*\*\*\* 1.2.2 Prueba canónica de Anderson

.

. ivreg2 ha\_nchs $X (D=of\_op)

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 5.27

Prob > F = 0.0000

Total (centered) SS = 1230.139487 Centered R2 = 0.0560

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0663

Residual SS = 1161.282359 Root MSE = .5388

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2560193 .1230562 2.08 0.037 .0148335 .4972051

personas | -.0092382 .0052317 -1.77 0.077 -.0194921 .0010157

orden\_n | -.0638571 .0284765 -2.24 0.025 -.1196701 -.0080442

ocupado\_jefe | -.036992 .0243693 -1.52 0.129 -.0847548 .0107709

educa\_jefe | .0060156 .0024312 2.47 0.013 .0012505 .0107808

ingresos\_h~e | .0001607 .0001074 1.50 0.134 -.0000497 .0003711

\_cons | -.0871872 .07087 -1.23 0.219 -.2260898 .0517155

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 77.954

Chi-sq(1) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 79.365

Stock-Yogo weak ID test critical values: 10% maximal IV size 16.38

15% maximal IV size 8.96

20% maximal IV size 6.66

25% maximal IV size 5.53

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.000

(equation exactly identified)

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: of\_op

------------------------------------------------------------------------------

.

. \*\*\*\* 1.2.3 Prueba Cragg-Donald

.

. ivreg2 ha\_nchs $X (D=of\_op), first

First-stage regressions

-----------------------

First-stage regression of D:

OLS estimation

--------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 24.03

Prob > F = 0.0000

Total (centered) SS = 999.424 Centered R2 = 0.0349

Total (uncentered) SS = 1952 Uncentered R2 = 0.5058

Residual SS = 964.5908114 Root MSE = .4915

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.021453 .004128 -5.20 0.000 -.0295461 -.0133598

orden\_n | .0934295 .0239443 3.90 0.000 .0464852 .1403738

ocupado\_jefe | .0772656 .0203094 3.80 0.000 .0374479 .1170834

educa\_jefe | -.0057974 .002117 -2.74 0.006 -.009948 -.0016469

ingresos\_h~e | .0003029 .000092 3.29 0.001 .0001226 .0004832

of\_op | .0336993 .0037827 8.91 0.000 .026283 .0411156

\_cons | .3567943 .0419026 8.51 0.000 .2746417 .4389468

------------------------------------------------------------------------------

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe of\_op

------------------------------------------------------------------------------

Partial R-squared of excluded instruments: 0.0195

Test of excluded instruments:

F( 1, 3993) = 79.36

Prob > F = 0.0000

Summary results for first-stage regressions

-------------------------------------------

Variable | Shea Partial R2 | Partial R2 | F( 1, 3993) P-value

D | 0.0195 | 0.0195 | 79.36 0.0000

Underidentification tests

Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)

Ha: matrix has rank=K1 (identified)

Anderson canon. corr. N\*CCEV LM statistic Chi-sq(1)=77.95 P-val=0.0000

Cragg-Donald N\*CDEV Wald statistic Chi-sq(1)=79.50 P-val=0.0000

Weak identification test

Ho: equation is weakly identified

Cragg-Donald Wald F-statistic 79.36

See main output for Cragg-Donald weak id test critical values

Weak-instrument-robust inference

Tests of joint significance of endogenous regressors B1 in main equation

Ho: B1=0 and overidentifying restrictions are valid

Anderson-Rubin Wald test F(1,3993)=4.11 P-val=0.0427

Anderson-Rubin Wald test Chi-sq(1)=4.12 P-val=0.0425

Stock-Wright LM S statistic Chi-sq(1)=4.11 P-val=0.0426

Number of observations N = 4000

Number of regressors K = 7

Number of instruments L = 7

Number of excluded instruments L1 = 1

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 5.27

Prob > F = 0.0000

Total (centered) SS = 1230.139487 Centered R2 = 0.0560

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0663

Residual SS = 1161.282359 Root MSE = .5388

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2560193 .1230562 2.08 0.037 .0148335 .4972051

personas | -.0092382 .0052317 -1.77 0.077 -.0194921 .0010157

orden\_n | -.0638571 .0284765 -2.24 0.025 -.1196701 -.0080442

ocupado\_jefe | -.036992 .0243693 -1.52 0.129 -.0847548 .0107709

educa\_jefe | .0060156 .0024312 2.47 0.013 .0012505 .0107808

ingresos\_h~e | .0001607 .0001074 1.50 0.134 -.0000497 .0003711

\_cons | -.0871872 .07087 -1.23 0.219 -.2260898 .0517155

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 77.954

Chi-sq(1) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 79.365

Stock-Yogo weak ID test critical values: 10% maximal IV size 16.38

15% maximal IV size 8.96

20% maximal IV size 6.66

25% maximal IV size 5.53

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.000

(equation exactly identified)

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: of\_op

------------------------------------------------------------------------------

.

. \*\*\*\* 1.2.4 Prueba de Stock y Yoko

.

. ivreg2 ha\_nchs $X (D=of\_op), first

First-stage regressions

-----------------------

First-stage regression of D:

OLS estimation

--------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 24.03

Prob > F = 0.0000

Total (centered) SS = 999.424 Centered R2 = 0.0349

Total (uncentered) SS = 1952 Uncentered R2 = 0.5058

Residual SS = 964.5908114 Root MSE = .4915

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.021453 .004128 -5.20 0.000 -.0295461 -.0133598

orden\_n | .0934295 .0239443 3.90 0.000 .0464852 .1403738

ocupado\_jefe | .0772656 .0203094 3.80 0.000 .0374479 .1170834

educa\_jefe | -.0057974 .002117 -2.74 0.006 -.009948 -.0016469

ingresos\_h~e | .0003029 .000092 3.29 0.001 .0001226 .0004832

of\_op | .0336993 .0037827 8.91 0.000 .026283 .0411156

\_cons | .3567943 .0419026 8.51 0.000 .2746417 .4389468

------------------------------------------------------------------------------

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe of\_op

------------------------------------------------------------------------------

Partial R-squared of excluded instruments: 0.0195

Test of excluded instruments:

F( 1, 3993) = 79.36

Prob > F = 0.0000

Summary results for first-stage regressions

-------------------------------------------

Variable | Shea Partial R2 | Partial R2 | F( 1, 3993) P-value

D | 0.0195 | 0.0195 | 79.36 0.0000

Underidentification tests

Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)

Ha: matrix has rank=K1 (identified)

Anderson canon. corr. N\*CCEV LM statistic Chi-sq(1)=77.95 P-val=0.0000

Cragg-Donald N\*CDEV Wald statistic Chi-sq(1)=79.50 P-val=0.0000

Weak identification test

Ho: equation is weakly identified

Cragg-Donald Wald F-statistic 79.36

See main output for Cragg-Donald weak id test critical values

Weak-instrument-robust inference

Tests of joint significance of endogenous regressors B1 in main equation

Ho: B1=0 and overidentifying restrictions are valid

Anderson-Rubin Wald test F(1,3993)=4.11 P-val=0.0427

Anderson-Rubin Wald test Chi-sq(1)=4.12 P-val=0.0425

Stock-Wright LM S statistic Chi-sq(1)=4.11 P-val=0.0426

Number of observations N = 4000

Number of regressors K = 7

Number of instruments L = 7

Number of excluded instruments L1 = 1

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 5.27

Prob > F = 0.0000

Total (centered) SS = 1230.139487 Centered R2 = 0.0560

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0663

Residual SS = 1161.282359 Root MSE = .5388

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2560193 .1230562 2.08 0.037 .0148335 .4972051

personas | -.0092382 .0052317 -1.77 0.077 -.0194921 .0010157

orden\_n | -.0638571 .0284765 -2.24 0.025 -.1196701 -.0080442

ocupado\_jefe | -.036992 .0243693 -1.52 0.129 -.0847548 .0107709

educa\_jefe | .0060156 .0024312 2.47 0.013 .0012505 .0107808

ingresos\_h~e | .0001607 .0001074 1.50 0.134 -.0000497 .0003711

\_cons | -.0871872 .07087 -1.23 0.219 -.2260898 .0517155

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 77.954

Chi-sq(1) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 79.365

Stock-Yogo weak ID test critical values: 10% maximal IV size 16.38

15% maximal IV size 8.96

20% maximal IV size 6.66

25% maximal IV size 5.53

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.000

(equation exactly identified)

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: of\_op

------------------------------------------------------------------------------

.

. \*\*\*\* Ahora realizamos las mismas pruebas utilizando los dos instrumentos:

.

. \*\*\*\* 1.3 Distancia y oficinas operadoras:

.

. \*\*\*\* 1.3.1 Regresión MCO

.

. reg D of\_op distancia

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 2, 3997) = 64.00

Model | 31.0125359 2 15.506268 Prob > F = 0.0000

Residual | 968.411464 3997 .242284579 R-squared = 0.0310

-------------+------------------------------ Adj R-squared = 0.0305

Total | 999.424 3999 .24991848 Root MSE = .49222

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

of\_op | .033219 .0037878 8.77 0.000 .0257928 .0406453

distancia | -.0000486 6.97e-06 -6.98 0.000 -.0000623 -.000035

\_cons | .4983069 .0191481 26.02 0.000 .4607659 .5358478

------------------------------------------------------------------------------

.

. reg D $X of\_op distancia

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 7, 3992) = 27.95

Model | 46.689176 7 6.66988229 Prob > F = 0.0000

Residual | 952.734824 3992 .238661028 R-squared = 0.0467

-------------+------------------------------ Adj R-squared = 0.0450

Total | 999.424 3999 .24991848 Root MSE = .48853

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.0213117 .0041031 -5.19 0.000 -.029356 -.0132673

orden\_n | .0972524 .0238059 4.09 0.000 .0505796 .1439253

ocupado\_jefe | .0753753 .0201885 3.73 0.000 .0357946 .1149561

educa\_jefe | -.0057376 .0021042 -2.73 0.006 -.0098631 -.0016121

ingresos\_h~e | .0003076 .0000914 3.37 0.001 .0001284 .0004868

of\_op | .0332178 .0037605 8.83 0.000 .0258451 .0405905

distancia | -.0000487 6.92e-06 -7.05 0.000 -.0000623 -.0000352

\_cons | .4527444 .0438179 10.33 0.000 .3668368 .538652

------------------------------------------------------------------------------

.

. test distancia=of\_op=0

( 1) - of\_op + distancia = 0

( 2) distancia = 0

F( 2, 3992) = 65.00

Prob > F = 0.0000

.

. \*\*\*\* 1.3.2 Prueba canónica de Anderson

.

. ivreg2 ha\_nchs $X (D=of\_op distancia)

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 5.33

Prob > F = 0.0000

Total (centered) SS = 1230.139487 Centered R2 = 0.0548

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0651

Residual SS = 1162.702759 Root MSE = .5391

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2102978 .0967891 2.17 0.030 .0205946 .400001

personas | -.0102136 .0049767 -2.05 0.040 -.0199678 -.0004595

orden\_n | -.0597519 .0276622 -2.16 0.031 -.1139689 -.0055349

ocupado\_jefe | -.0333079 .0236003 -1.41 0.158 -.0795636 .0129479

educa\_jefe | .0057464 .0023911 2.40 0.016 .00106 .0104329

ingresos\_h~e | .0001744 .000105 1.66 0.097 -.0000313 .0003801

\_cons | -.0667076 .0621803 -1.07 0.283 -.1885789 .0551636

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 126.161

Chi-sq(2) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 65.005

Stock-Yogo weak ID test critical values: 10% maximal IV size 19.93

15% maximal IV size 11.59

20% maximal IV size 8.75

25% maximal IV size 7.25

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.361

Chi-sq(1) P-val = 0.5480

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: of\_op distancia

------------------------------------------------------------------------------

.

. \*\*\*\* 1.3.3 Prueba Cragg-Donald

.

. ivreg2 ha\_nchs $X (D=of\_op distancia), first

First-stage regressions

-----------------------

First-stage regression of D:

OLS estimation

--------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 7, 3992) = 27.95

Prob > F = 0.0000

Total (centered) SS = 999.424 Centered R2 = 0.0467

Total (uncentered) SS = 1952 Uncentered R2 = 0.5119

Residual SS = 952.734824 Root MSE = .4885

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.0213117 .0041031 -5.19 0.000 -.029356 -.0132673

orden\_n | .0972524 .0238059 4.09 0.000 .0505796 .1439253

ocupado\_jefe | .0753753 .0201885 3.73 0.000 .0357946 .1149561

educa\_jefe | -.0057376 .0021042 -2.73 0.006 -.0098631 -.0016121

ingresos\_h~e | .0003076 .0000914 3.37 0.001 .0001284 .0004868

of\_op | .0332178 .0037605 8.83 0.000 .0258451 .0405905

distancia | -.0000487 6.92e-06 -7.05 0.000 -.0000623 -.0000352

\_cons | .4527444 .0438179 10.33 0.000 .3668368 .538652

------------------------------------------------------------------------------

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe of\_op distancia

------------------------------------------------------------------------------

Partial R-squared of excluded instruments: 0.0315

Test of excluded instruments:

F( 2, 3992) = 65.00

Prob > F = 0.0000

Summary results for first-stage regressions

-------------------------------------------

Variable | Shea Partial R2 | Partial R2 | F( 2, 3992) P-value

D | 0.0315 | 0.0315 | 65.00 0.0000

Underidentification tests

Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)

Ha: matrix has rank=K1 (identified)

Anderson canon. corr. N\*CCEV LM statistic Chi-sq(2)=126.16 P-val=0.0000

Cragg-Donald N\*CDEV Wald statistic Chi-sq(2)=130.27 P-val=0.0000

Weak identification test

Ho: equation is weakly identified

Cragg-Donald Wald F-statistic 65.00

See main output for Cragg-Donald weak id test critical values

Weak-instrument-robust inference

Tests of joint significance of endogenous regressors B1 in main equation

Ho: B1=0 and overidentifying restrictions are valid

Anderson-Rubin Wald test F(2,3992)=2.42 P-val=0.0895

Anderson-Rubin Wald test Chi-sq(2)=4.84 P-val=0.0889

Stock-Wright LM S statistic Chi-sq(2)=4.83 P-val=0.0892

Number of observations N = 4000

Number of regressors K = 7

Number of instruments L = 8

Number of excluded instruments L1 = 2

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 5.33

Prob > F = 0.0000

Total (centered) SS = 1230.139487 Centered R2 = 0.0548

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0651

Residual SS = 1162.702759 Root MSE = .5391

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2102978 .0967891 2.17 0.030 .0205946 .400001

personas | -.0102136 .0049767 -2.05 0.040 -.0199678 -.0004595

orden\_n | -.0597519 .0276622 -2.16 0.031 -.1139689 -.0055349

ocupado\_jefe | -.0333079 .0236003 -1.41 0.158 -.0795636 .0129479

educa\_jefe | .0057464 .0023911 2.40 0.016 .00106 .0104329

ingresos\_h~e | .0001744 .000105 1.66 0.097 -.0000313 .0003801

\_cons | -.0667076 .0621803 -1.07 0.283 -.1885789 .0551636

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 126.161

Chi-sq(2) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 65.005

Stock-Yogo weak ID test critical values: 10% maximal IV size 19.93

15% maximal IV size 11.59

20% maximal IV size 8.75

25% maximal IV size 7.25

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.361

Chi-sq(1) P-val = 0.5480

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: of\_op distancia

------------------------------------------------------------------------------

.

. \*\*\*\* 1.3.4 Prueba de Stock y Yoko

.

. ivreg2 ha\_nchs $X (D=of\_op distancia), first

First-stage regressions

-----------------------

First-stage regression of D:

OLS estimation

--------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 7, 3992) = 27.95

Prob > F = 0.0000

Total (centered) SS = 999.424 Centered R2 = 0.0467

Total (uncentered) SS = 1952 Uncentered R2 = 0.5119

Residual SS = 952.734824 Root MSE = .4885

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.0213117 .0041031 -5.19 0.000 -.029356 -.0132673

orden\_n | .0972524 .0238059 4.09 0.000 .0505796 .1439253

ocupado\_jefe | .0753753 .0201885 3.73 0.000 .0357946 .1149561

educa\_jefe | -.0057376 .0021042 -2.73 0.006 -.0098631 -.0016121

ingresos\_h~e | .0003076 .0000914 3.37 0.001 .0001284 .0004868

of\_op | .0332178 .0037605 8.83 0.000 .0258451 .0405905

distancia | -.0000487 6.92e-06 -7.05 0.000 -.0000623 -.0000352

\_cons | .4527444 .0438179 10.33 0.000 .3668368 .538652

------------------------------------------------------------------------------

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe of\_op distancia

------------------------------------------------------------------------------

Partial R-squared of excluded instruments: 0.0315

Test of excluded instruments:

F( 2, 3992) = 65.00

Prob > F = 0.0000

Summary results for first-stage regressions

-------------------------------------------

Variable | Shea Partial R2 | Partial R2 | F( 2, 3992) P-value

D | 0.0315 | 0.0315 | 65.00 0.0000

Underidentification tests

Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)

Ha: matrix has rank=K1 (identified)

Anderson canon. corr. N\*CCEV LM statistic Chi-sq(2)=126.16 P-val=0.0000

Cragg-Donald N\*CDEV Wald statistic Chi-sq(2)=130.27 P-val=0.0000

Weak identification test

Ho: equation is weakly identified

Cragg-Donald Wald F-statistic 65.00

See main output for Cragg-Donald weak id test critical values

Weak-instrument-robust inference

Tests of joint significance of endogenous regressors B1 in main equation

Ho: B1=0 and overidentifying restrictions are valid

Anderson-Rubin Wald test F(2,3992)=2.42 P-val=0.0895

Anderson-Rubin Wald test Chi-sq(2)=4.84 P-val=0.0889

Stock-Wright LM S statistic Chi-sq(2)=4.83 P-val=0.0892

Number of observations N = 4000

Number of regressors K = 7

Number of instruments L = 8

Number of excluded instruments L1 = 2

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 5.33

Prob > F = 0.0000

Total (centered) SS = 1230.139487 Centered R2 = 0.0548

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0651

Residual SS = 1162.702759 Root MSE = .5391

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2102978 .0967891 2.17 0.030 .0205946 .400001

personas | -.0102136 .0049767 -2.05 0.040 -.0199678 -.0004595

orden\_n | -.0597519 .0276622 -2.16 0.031 -.1139689 -.0055349

ocupado\_jefe | -.0333079 .0236003 -1.41 0.158 -.0795636 .0129479

educa\_jefe | .0057464 .0023911 2.40 0.016 .00106 .0104329

ingresos\_h~e | .0001744 .000105 1.66 0.097 -.0000313 .0003801

\_cons | -.0667076 .0621803 -1.07 0.283 -.1885789 .0551636

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 126.161

Chi-sq(2) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 65.005

Stock-Yogo weak ID test critical values: 10% maximal IV size 19.93

15% maximal IV size 11.59

20% maximal IV size 8.75

25% maximal IV size 7.25

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.361

Chi-sq(1) P-val = 0.5480

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: of\_op distancia

------------------------------------------------------------------------------

.

. \*\*\*\* Podemos concluir, por lo tanto, que los instrumentos son relevantes.

.

.

. \*------------------\*

. \* 2. Exogeneidad \*

. \*------------------\*

.

. \*\*\*\* Al igual que probamos la relevancia de los instrumentos, la exogeneidad la probaremos

. \*\*\*\* en cada uno de los instrumentos y después utilizando los dos.

.

. \*\*\*\* 2.1 Distancia

.

. \*\*\*\* 2.1.1 Regresión de errores en función del instrumento.

.

. \*\*\*\* Lo que vamos a hacer es predecir los errores de MCO y verificar que estos no estén relacionados

. \*\*\*\* con el instrumento

.

. reg ha\_nchs D $X

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 6, 3993) = 39.49

Model | 68.9063178 6 11.4843863 Prob > F = 0.0000

Residual | 1161.23317 3993 .290817222 R-squared = 0.0560

-------------+------------------------------ Adj R-squared = 0.0546

Total | 1230.13949 3999 .307611775 Root MSE = .53927

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2489481 .0171935 14.48 0.000 .2152392 .282657

personas | -.0093891 .004544 -2.07 0.039 -.018298 -.0004802

orden\_n | -.0632222 .0263134 -2.40 0.016 -.1148111 -.0116333

ocupado\_jefe | -.0364222 .0223229 -1.63 0.103 -.0801875 .0073431

educa\_jefe | .005974 .002325 2.57 0.010 .0014157 .0105322

ingresos\_h~e | .0001628 .000101 1.61 0.107 -.0000353 .0003609

\_cons | -.0840199 .0452452 -1.86 0.063 -.1727257 .004686

------------------------------------------------------------------------------

.

. predict uhat, residuals

.

. reg uhat distancia

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 1, 3998) = 0.52

Model | .149670856 1 .149670856 Prob > F = 0.4729

Residual | 1161.0835 3998 .290416083 R-squared = 0.0001

-------------+------------------------------ Adj R-squared = -0.0001

Total | 1161.23317 3999 .290380888 Root MSE = .5389

------------------------------------------------------------------------------

uhat | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

distancia | 5.47e-06 7.63e-06 0.72 0.473 -9.48e-06 .0000204

\_cons | -.0110969 .0176505 -0.63 0.530 -.0457017 .023508

------------------------------------------------------------------------------

.

. \*\*\*\* Vemos que la variable distancia no es significativa y tampoco lo es el modelo.

. \*\*\*\* Para utilizar las otras pruebas de exogeneidad es necesario que el número

. \*\*\*\* de instrumentso exceda el número de variables a instrumentar. Estas pruebas

. \*\*\*\* serán llevadas a cabo cuando evaluemos la exogeneidad de la distancia y las

. \*\*\*\* oficinas operadoras de manera conjunta.

.

. \*\*\*\* 2.2 Oficinas operadoras

.

. \*\*\*\* 2.2.1 Regresión de errores en función del instrumento

.

. reg uhat of\_op

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 1, 3998) = 0.00

Model | .000958042 1 .000958042 Prob > F = 0.9542

Residual | 1161.23221 3998 .29045328 R-squared = 0.0000

-------------+------------------------------ Adj R-squared = -0.0002

Total | 1161.23317 3999 .290380888 Root MSE = .53894

------------------------------------------------------------------------------

uhat | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

of\_op | .0002381 .0041466 0.06 0.954 -.0078914 .0083677

\_cons | -.0006329 .0139299 -0.05 0.964 -.0279432 .0266775

------------------------------------------------------------------------------

.

. \*\*\*\* Vemos que of\_op aparentemente no predice bien los errores.

.

. \*\*\*\* 2.3 Distancia y Oficinas operadoras

.

. \*\*\*\* 2.2.1 Regresión de errores en función de los instrumentos

.

. reg uhat of\_op distancia

Source | SS df MS Number of obs = 4000

-------------+------------------------------ F( 2, 3997) = 0.26

Model | .151132554 2 .075566277 Prob > F = 0.7710

Residual | 1161.08204 3997 .290488376 R-squared = 0.0001

-------------+------------------------------ Adj R-squared = -0.0004

Total | 1161.23317 3999 .290380888 Root MSE = .53897

------------------------------------------------------------------------------

uhat | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

of\_op | .0002942 .0041475 0.07 0.943 -.0078373 .0084257

distancia | 5.48e-06 7.63e-06 0.72 0.472 -9.47e-06 .0000204

\_cons | -.0118993 .0209666 -0.57 0.570 -.0530055 .0292069

------------------------------------------------------------------------------

.

. \*\*\*\* 2.2.2 Prueba de Sargan

.

. ivreg2 ha\_nchs $X (D=of\_op distancia)

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 5.33

Prob > F = 0.0000

Total (centered) SS = 1230.139487 Centered R2 = 0.0548

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0651

Residual SS = 1162.702759 Root MSE = .5391

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2102978 .0967891 2.17 0.030 .0205946 .400001

personas | -.0102136 .0049767 -2.05 0.040 -.0199678 -.0004595

orden\_n | -.0597519 .0276622 -2.16 0.031 -.1139689 -.0055349

ocupado\_jefe | -.0333079 .0236003 -1.41 0.158 -.0795636 .0129479

educa\_jefe | .0057464 .0023911 2.40 0.016 .00106 .0104329

ingresos\_h~e | .0001744 .000105 1.66 0.097 -.0000313 .0003801

\_cons | -.0667076 .0621803 -1.07 0.283 -.1885789 .0551636

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 126.161

Chi-sq(2) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 65.005

Stock-Yogo weak ID test critical values: 10% maximal IV size 19.93

15% maximal IV size 11.59

20% maximal IV size 8.75

25% maximal IV size 7.25

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.361

Chi-sq(1) P-val = 0.5480

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: of\_op distancia

------------------------------------------------------------------------------

.

. \*\*\*\* La hipótesis nula de esta prueba es que los instrumentos son exógenos. Rechazarla nos llevaría a

> dudar

. \*\*\*\* sobre la validez de los instrumentos. En este caso no podemos rechazar esta prueba.

.

. \*\*\*\* 2.2.3 Sargan y Basmann

.

. overid

Tests of overidentifying restrictions:

Sargan N\*R-sq test 0.361 Chi-sq(1) P-value = 0.5480

Basmann test 0.360 Chi-sq(1) P-value = 0.5484

.

. \*\*\*\* Utilizando el comando "overid" luego de llevar a cabo una estimación por variables instrumentales

. \*\*\*\* se lleva a cabo la prueba de Sargan y la de Basmann. La prueba de Sargan ya la vimos, la de Basma

> nn

. \*\*\*\* es una prueba análoga. Sin embargo, el estadístico de prueba se construye de manera diferente.

.

. \*\*\*\*2.2.4 Prueba "J".

.

. \*\*\*\* La prueba "J" es análoga a la prueba de Sargan pero se utiliza cuando la estimación se hace media

> nte

. \*\*\*\* el método generalizado de momentos:

.

. ivreg2 ha\_nchs $X (D=of\_op distancia), gmm2s robust

2-Step GMM estimation

---------------------

Estimates efficient for arbitrary heteroskedasticity

Statistics robust to heteroskedasticity

Number of obs = 4000

F( 6, 3993) = 5.55

Prob > F = 0.0000

Total (centered) SS = 1230.139487 Centered R2 = 0.0546

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0649

Residual SS = 1163.013918 Root MSE = .5392

------------------------------------------------------------------------------

| Robust

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2064098 .1010242 2.04 0.041 .008406 .4044136

personas | -.0103566 .0046438 -2.23 0.026 -.0194583 -.0012549

orden\_n | -.0592206 .0284924 -2.08 0.038 -.1150647 -.0033765

ocupado\_jefe | -.0324303 .0229689 -1.41 0.158 -.0774486 .012588

educa\_jefe | .005722 .0023758 2.41 0.016 .0010655 .0103785

ingresos\_h~e | .0001744 .0001107 1.58 0.115 -.0000426 .0003914

\_cons | -.0649516 .0621645 -1.04 0.296 -.1867917 .0568885

------------------------------------------------------------------------------

Underidentification test (Kleibergen-Paap rk LM statistic): 125.143

Chi-sq(2) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Kleibergen-Paap rk Wald F statistic): 69.694

Stock-Yogo weak ID test critical values: 10% maximal IV size 19.93

15% maximal IV size 11.59

20% maximal IV size 8.75

25% maximal IV size 7.25

Source: Stock-Yogo (2005). Reproduced by permission.

NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.

------------------------------------------------------------------------------

Hansen J statistic (overidentification test of all instruments): 0.365

Chi-sq(1) P-val = 0.5456

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: of\_op distancia

------------------------------------------------------------------------------

.

.

. \*----------------------------\*

. \* 3. Estimación por MC2E \*

. \*----------------------------\*

.

. \*\*\*\* Dado que ya probamos que los instrumentos que tenemos pueden ser utilizados, procedemos

. \*\*\*\* a realizar la estimación mediante MCO en dos etapas.

.

. \*\*\*\* 3.1 Primera etapa, predicción de la variable endógena (D) utilizando los

. \*\*\*\* regresores exógenos y los instrumentos (distancia y of\_op):

.

. drop uhat

.

. probit D $X distancia of\_op

Iteration 0: log likelihood = -2771.4366

Iteration 1: log likelihood = -2675.7646

Iteration 2: log likelihood = -2675.6035

Iteration 3: log likelihood = -2675.6035

Probit regression Number of obs = 4000

LR chi2(7) = 191.67

Prob > chi2 = 0.0000

Log likelihood = -2675.6035 Pseudo R2 = 0.0346

------------------------------------------------------------------------------

D | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.056137 .0108297 -5.18 0.000 -.0773628 -.0349112

orden\_n | .2556066 .0626281 4.08 0.000 .1328577 .3783555

ocupado\_jefe | .1947698 .0528018 3.69 0.000 .0912802 .2982595

educa\_jefe | -.015059 .0054907 -2.74 0.006 -.0258205 -.0042975

ingresos\_h~e | .0008236 .0002426 3.40 0.001 .0003482 .0012991

distancia | -.0001267 .0000181 -7.00 0.000 -.0001622 -.0000912

of\_op | .0867556 .0099521 8.72 0.000 .0672497 .1062614

\_cons | -.1219782 .1145348 -1.06 0.287 -.3464623 .1025059

------------------------------------------------------------------------------

.

. predict trat\_hat

(option pr assumed; Pr(D))

.

.

. \*\*\*\* 3.2 Segunda etapa, regresión donde la variable dependiente es la talla para la edad y

. \*\*\*\* se utilizan las variables predichas por la anterior regresión (trata\_hat) para instrumentar

. \*\*\*\* la participación en el programa:

.

. ivreg2 ha\_nchs $X (D=trat\_hat), first

First-stage regressions

-----------------------

First-stage regression of D:

OLS estimation

--------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 32.69

Prob > F = 0.0000

Total (centered) SS = 999.424 Centered R2 = 0.0468

Total (uncentered) SS = 1952 Uncentered R2 = 0.5120

Residual SS = 952.6320316 Root MSE = .4884

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | .0000419 .0045089 0.01 0.993 -.0087981 .0088819

orden\_n | -.000604 .0250734 -0.02 0.981 -.0497619 .0485539

ocupado\_jefe | .0006032 .0213597 0.03 0.977 -.0412737 .0424801

educa\_jefe | -2.57e-06 .002166 -0.00 0.999 -.0042491 .004244

ingresos\_h~e | -3.32e-07 .0000951 -0.00 0.997 -.0001868 .0001861

trat\_hat | 1.000707 .087604 11.42 0.000 .8289546 1.17246

\_cons | -.0004202 .0563133 -0.01 0.994 -.1108256 .1099852

------------------------------------------------------------------------------

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe trat\_hat

------------------------------------------------------------------------------

Partial R-squared of excluded instruments: 0.0316

Test of excluded instruments:

F( 1, 3993) = 130.49

Prob > F = 0.0000

Summary results for first-stage regressions

-------------------------------------------

Variable | Shea Partial R2 | Partial R2 | F( 1, 3993) P-value

D | 0.0316 | 0.0316 | 130.49 0.0000

Underidentification tests

Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)

Ha: matrix has rank=K1 (identified)

Anderson canon. corr. N\*CCEV LM statistic Chi-sq(1)=126.58 P-val=0.0000

Cragg-Donald N\*CDEV Wald statistic Chi-sq(1)=130.72 P-val=0.0000

Weak identification test

Ho: equation is weakly identified

Cragg-Donald Wald F-statistic 130.49

See main output for Cragg-Donald weak id test critical values

Weak-instrument-robust inference

Tests of joint significance of endogenous regressors B1 in main equation

Ho: B1=0 and overidentifying restrictions are valid

Anderson-Rubin Wald test F(1,3993)=4.92 P-val=0.0266

Anderson-Rubin Wald test Chi-sq(1)=4.93 P-val=0.0264

Stock-Wright LM S statistic Chi-sq(1)=4.92 P-val=0.0265

Number of observations N = 4000

Number of regressors K = 7

Number of instruments L = 7

Number of excluded instruments L1 = 1

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 5.41

Prob > F = 0.0000

Total (centered) SS = 1230.139487 Centered R2 = 0.0553

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0656

Residual SS = 1162.070617 Root MSE = .539

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2197716 .0966029 2.27 0.023 .0304333 .4091098

personas | -.0100115 .0049739 -2.01 0.044 -.0197602 -.0002628

orden\_n | -.0606025 .0276502 -2.19 0.028 -.1147959 -.0064091

ocupado\_jefe | -.0340712 .0235896 -1.44 0.149 -.0803061 .0121636

educa\_jefe | .0058022 .0023902 2.43 0.015 .0011175 .010487

ingresos\_h~e | .0001716 .0001049 1.64 0.102 -.0000341 .0003772

\_cons | -.0709511 .0621135 -1.14 0.253 -.1926914 .0507892

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 126.579

Chi-sq(1) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 130.487

Stock-Yogo weak ID test critical values: 10% maximal IV size 16.38

15% maximal IV size 8.96

20% maximal IV size 6.66

25% maximal IV size 5.53

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.000

(equation exactly identified)

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: trat\_hat

------------------------------------------------------------------------------

.

. \*\*\*\* Para llevar a cabo la estimación de manera directa con el comando "ivreg2" instrumentando el trat

> amiento con la distancia y el

. \*\*\*\* número de oficinas operadoras en el municipio:

.

. ivreg2 ha\_nchs $X (D=distancia of\_op), first

First-stage regressions

-----------------------

First-stage regression of D:

OLS estimation

--------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 7, 3992) = 27.95

Prob > F = 0.0000

Total (centered) SS = 999.424 Centered R2 = 0.0467

Total (uncentered) SS = 1952 Uncentered R2 = 0.5119

Residual SS = 952.734824 Root MSE = .4885

------------------------------------------------------------------------------

D | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

personas | -.0213117 .0041031 -5.19 0.000 -.029356 -.0132673

orden\_n | .0972524 .0238059 4.09 0.000 .0505796 .1439253

ocupado\_jefe | .0753753 .0201885 3.73 0.000 .0357946 .1149561

educa\_jefe | -.0057376 .0021042 -2.73 0.006 -.0098631 -.0016121

ingresos\_h~e | .0003076 .0000914 3.37 0.001 .0001284 .0004868

distancia | -.0000487 6.92e-06 -7.05 0.000 -.0000623 -.0000352

of\_op | .0332178 .0037605 8.83 0.000 .0258451 .0405905

\_cons | .4527444 .0438179 10.33 0.000 .3668368 .538652

------------------------------------------------------------------------------

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe distancia of\_op

------------------------------------------------------------------------------

Partial R-squared of excluded instruments: 0.0315

Test of excluded instruments:

F( 2, 3992) = 65.00

Prob > F = 0.0000

Summary results for first-stage regressions

-------------------------------------------

Variable | Shea Partial R2 | Partial R2 | F( 2, 3992) P-value

D | 0.0315 | 0.0315 | 65.00 0.0000

Underidentification tests

Ho: matrix of reduced form coefficients has rank=K1-1 (underidentified)

Ha: matrix has rank=K1 (identified)

Anderson canon. corr. N\*CCEV LM statistic Chi-sq(2)=126.16 P-val=0.0000

Cragg-Donald N\*CDEV Wald statistic Chi-sq(2)=130.27 P-val=0.0000

Weak identification test

Ho: equation is weakly identified

Cragg-Donald Wald F-statistic 65.00

See main output for Cragg-Donald weak id test critical values

Weak-instrument-robust inference

Tests of joint significance of endogenous regressors B1 in main equation

Ho: B1=0 and overidentifying restrictions are valid

Anderson-Rubin Wald test F(2,3992)=2.42 P-val=0.0895

Anderson-Rubin Wald test Chi-sq(2)=4.84 P-val=0.0889

Stock-Wright LM S statistic Chi-sq(2)=4.83 P-val=0.0892

Number of observations N = 4000

Number of regressors K = 7

Number of instruments L = 8

Number of excluded instruments L1 = 2

IV (2SLS) estimation

--------------------

Estimates efficient for homoskedasticity only

Statistics consistent for homoskedasticity only

Number of obs = 4000

F( 6, 3993) = 5.33

Prob > F = 0.0000

Total (centered) SS = 1230.139487 Centered R2 = 0.0548

Total (uncentered) SS = 1243.678874 Uncentered R2 = 0.0651

Residual SS = 1162.702759 Root MSE = .5391

------------------------------------------------------------------------------

ha\_nchs | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

D | .2102978 .0967891 2.17 0.030 .0205946 .400001

personas | -.0102136 .0049767 -2.05 0.040 -.0199678 -.0004595

orden\_n | -.0597519 .0276622 -2.16 0.031 -.1139689 -.0055349

ocupado\_jefe | -.0333079 .0236003 -1.41 0.158 -.0795636 .0129479

educa\_jefe | .0057464 .0023911 2.40 0.016 .00106 .0104329

ingresos\_h~e | .0001744 .000105 1.66 0.097 -.0000313 .0003801

\_cons | -.0667076 .0621803 -1.07 0.283 -.1885789 .0551636

------------------------------------------------------------------------------

Underidentification test (Anderson canon. corr. LM statistic): 126.161

Chi-sq(2) P-val = 0.0000

------------------------------------------------------------------------------

Weak identification test (Cragg-Donald Wald F statistic): 65.005

Stock-Yogo weak ID test critical values: 10% maximal IV size 19.93

15% maximal IV size 11.59

20% maximal IV size 8.75

25% maximal IV size 7.25

Source: Stock-Yogo (2005). Reproduced by permission.

------------------------------------------------------------------------------

Sargan statistic (overidentification test of all instruments): 0.361

Chi-sq(1) P-val = 0.5480

------------------------------------------------------------------------------

Instrumented: D

Included instruments: personas orden\_n ocupado\_jefe educa\_jefe

ingresos\_hogar\_jefe

Excluded instruments: distancia of\_op

------------------------------------------------------------------------------

.

. \*\*\*\* Utilizando mínimos cuadrados en dos etapas vemos que el efecto del tratamiento se disminuye significa

> tivamente.

. \*\*\*\* Mediante MCO el efecto del tratamiento en la talla para la edad era de 0.24 desviaciones estándar. Si

> n embargo,

. \*\*\*\* mediante MC2E pudimos confirmar que en realidad el efecto es 0.21.