## Problem set 5

Duarte Baracat, Agustín; Oviedo, Nicolás; Sánchez, Guido; Sosa, Juan Bautista

Economía Aplicada 2020 - UdeSA

## **Ejercicio 1**

Para empezar, reproducimos la Tabla 1 del trabajo de Card (1993). Hay que aclarar que la primera columna no pudo ser replicada dado que los datos menesteres no estaban contenidos en la base con la que trabajamos. Por otra parte, en la base de datos contamos con 9 indicadores regionales, pero el autor presenta sólo cuatro regiones. En consecuencia, hemos debido inferir cuáles variables corresponden a qué regiones. Nos manejamos de aquí en más con las siguientes definiciones:

Sur = reg661 + reg662 + reg665

Noreste = reg667 + reg669

Medio-Oeste = reg663 + reg664

Oeste = reg666 + reg668

Tabla 1

|  |  |  |
| --- | --- | --- |
|  | Subconjunto con valores válidos de educación | Subconjunto con valores válidos de educación y salario |
| Distribución de edad en 1966: |  |  |
| Edad 14-15 | 25.30 | 25.5 |
| Edad 16-17 | 23.83 | 24.1 |
| Edad 18-20 | 24.14 | 24.6 |
| Edad 21-24 | 26.74 | 25.8 |
|  |  |  |
| Distribución regional en 1966: |  |  |
| Sur | 40.91 | 20.7 |
| Noreste | 20.37 | 26 |
| Medio-oeste | 26.32 | 41.7 |
| Oeste | 12.40 | 11.9 |
| Vive en Zona Metropolitana en 1976 | 69.47 | 65 |
| Vive cerca de universidad | 67.81 | 68.2 |
| Vivía a los 14 años con madre y padre | 79.21 | 78.9 |
| Vivía a los 14 años solo con la madre | 10 | 10.1 |
| Educación promedio madre | 10.4 | 10.3 |
| Educación promedio padre | 10 | 10 |
| Porcentaje raza negra | 23 | 23 |
| Puntaje promedio de KWW | 33.5 | 33.5 |
| Educación promedio en 1976 | 13.2 | 13.3 |
| Vive en el sur 1976 | 40 | 40.3 |

## **Ejercicio 2**

A continuación, mostramos los resultados de las regresiones que hemos ejecutado para replicar la Tabla 2 del trabajo de Card. La presentamos de forma completa, mostrando las estimaciones y errores estándar de todos los parámetros.

Tabla 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) | (5) |
| VARIABLES | Log(wage) in 1976 | Log(wage) in 1976 | Log(wage) in 1976 | Log(wage) in 1976 | Log(wage) in 1976 |
|  |  |  |  |  |  |
| Education in 1976 | 0.0743\*\*\* | 0.0769\*\*\* | 0.0764\*\*\* | 0.0764\*\*\* | 0.0755\*\*\* |
|  | (0.0037) | (0.0037) | (0.0038) | (0.0038) | (0.0039) |
| Experience at 1976 | 0.0845\*\*\* | 0.0867\*\*\* | 0.0870\*\*\* | 0.0869\*\*\* | 0.0868\*\*\* |
|  | (0.0067) | (0.0068) | (0.0068) | (0.0068) | (0.0069) |
| Experience squared at 1976 | -0.0023\*\*\* | -0.0024\*\*\* | -0.0024\*\*\* | -0.0024\*\*\* | -0.0025\*\*\* |
|  | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| Black | -0.1896\*\*\* | -0.1844\*\*\* | -0.1820\*\*\* | -0.1817\*\*\* | -0.1767\*\*\* |
|  | (0.0182) | (0.0185) | (0.0193) | (0.0197) | (0.0199) |
| Lived in South in 1966 | -0.0975\*\*\* | 0.0822\* | 0.0833\* | 0.0819\* | 0.0811\* |
|  | (0.0159) | (0.0470) | (0.0471) | (0.0474) | (0.0474) |
| In smsa in 1966 |  | 0.1056\*\*\* | 0.1051\*\*\* | 0.1035\*\*\* | 0.1047\*\*\* |
|  |  | (0.0150) | (0.0150) | (0.0151) | (0.0151) |
| Region=1 in 1966 |  | 0.0569 | 0.0573 | 0.0514 | 0.0499 |
|  |  | (0.0511) | (0.0512) | (0.0516) | (0.0516) |
| Region=2 in 1966 |  | 0.1643\*\*\* | 0.1650\*\*\* | 0.1620\*\*\* | 0.1607\*\*\* |
|  |  | (0.0439) | (0.0440) | (0.0443) | (0.0444) |
| Region=3 in 1966 |  | 0.2061\*\*\* | 0.2068\*\*\* | 0.2032\*\*\* | 0.2016\*\*\* |
|  |  | (0.0430) | (0.0431) | (0.0434) | (0.0435) |
| Region=4 in 1966 |  | 0.1104\*\* | 0.1116\*\* | 0.1073\*\* | 0.1057\*\* |
|  |  | (0.0487) | (0.0487) | (0.0491) | (0.0493) |
| Region=5 in 1966 |  | -0.0193 | -0.0194 | -0.0192 | -0.0198 |
|  |  | (0.0265) | (0.0265) | (0.0266) | (0.0266) |
| Region=6 in 1966 = o, |  | - | - | - | - |
|  |  |  |  |  |  |
| Region=7 in 1966 |  | -0.0307 | -0.0308 | -0.0294 | -0.0288 |
|  |  | (0.0311) | (0.0311) | (0.0312) | (0.0312) |
| Region=8 in 1966 = o, |  | - | - | - | - |
|  |  |  |  |  |  |
| Region=9 in 1966 |  | 0.1852\*\*\* | 0.1855\*\*\* | 0.1834\*\*\* | 0.1844\*\*\* |
|  |  | (0.0469) | (0.0469) | (0.0471) | (0.0472) |
| In smsa in 1966 = o, |  | - | - | - | - |
|  |  |  |  |  |  |
| Mom-dad education class 1-9 |  |  | -0.0016 | 0.0032 | 0.0072 |
|  |  |  | (0.0032) | (0.0049) | (0.0051) |
| Mom and dad both > 12 yrs ed = o, |  |  |  | - | - |
|  |  |  |  |  |  |
| Mom and dad >=12 and not both exactly 12 |  |  |  | 0.0430 | 0.0400 |
|  |  |  |  | (0.0377) | (0.0378) |
| Mom=dad=12 |  |  |  | 0.0195 | 0.0132 |
|  |  |  |  | (0.0324) | (0.0325) |
| Mom >=12 and dad missing |  |  |  | 0.0132 | 0.0348 |
|  |  |  |  | (0.0367) | (0.0384) |
| Father >=12 and mom not in f1-f4 |  |  |  | 0.0326 | 0.0201 |
|  |  |  |  | (0.0301) | (0.0305) |
| Mom>=12 and dad nonmissing |  |  |  | 0.0213 | 0.0013 |
|  |  |  |  | (0.0252) | (0.0263) |
| Mom and dad both >=9 |  |  |  | 0.0154 | -0.0093 |
|  |  |  |  | (0.0334) | (0.0347) |
| Mom and dad both nonmissing |  |  |  | -0.0256 | -0.0544\*\* |
|  |  |  |  | (0.0201) | (0.0229) |
| 1 if lived with mom and dad at age 14 |  |  |  |  | 0.0535\*\* |
|  |  |  |  |  | (0.0249) |
| 1 if lived with single mom at age 14 |  |  |  |  | -0.0007 |
|  |  |  |  |  | (0.0310) |
| In smsa in 1976 | 0.1682\*\*\* |  |  |  |  |
|  | (0.0152) |  |  |  |  |
| Constant | 4.7119\*\*\* | 4.5612\*\*\* | 4.5742\*\*\* | 4.5416\*\*\* | 4.5018\*\*\* |
|  | (0.0703) | (0.0828) | (0.0866) | (0.0938) | (0.0962) |
|  |  |  |  |  |  |
| Observations | 3,010 | 3,010 | 3,010 | 3,010 | 3,010 |
| R-squared | 0.2836 | 0.2790 | 0.2790 | 0.2804 | 0.2818 |

## **Ejercicio 3**

A continuación, se replican los resultados de las regresiones que Card sintetiza en la Tabla 4.

|  |  |  |
| --- | --- | --- |
|  | Estimadores MCO | Estimadores VI |
| Especificación base | .07546678 | .167172 |
| Sueldos de 1978 | .07234716 | .14156502 |
| KWW | .05683242 | .20661071 |
| KWW con IQ |  | .13616421 |
| Proximidad a universidad pública |  | .12877128 |
| Proximidad a universidades con carreras de 2 y 4 años |  | .1278674 |
| Submuestra de 14 a 19 años | .07859929 | .15491332 |

## **Ejercicio 4**

Las variables instrumentales pueden servirnos para aislar la exogeneidad de una variable de interés, sin embargo es necesario que nuestro instrumento sea exógeno. Para ello debemos suponer la exogeneidad del instrumento ya que no es posible testearlo. Este es nuestro supuesto de inferencia y su justificación es lo que sostiene todo nuestro estudio. Sin embargo, una vez que establecemos como cierta la validez del instrumento, este supuesto nos permite testear la endogeneidad de nuestra variable de interés. Para ello podemos valernos del test de Hausman, lo que realizamos para las regresiones que acabamos de replicar. Como puede verse en la tabla 5, el test arroja un estadístico chi-cuadrado de -28,83, lo que nos indica que no podemos rechazar la hipótesis nula de que la diferencia entre el estimador de OLS y IV es no significativa. Esto implica que el estimador de OLS no es inconsistente dado nuestro supuesto de que IV es consistente. Sin embargo, no estamos seguros de que la implementación sea correcta.

## **Ejercicio 5**

El test de Sargan evalúa la validez de los instrumentos. El test supone que al menos un instrumento es válido, y evalúa si todos lo son. El estadístico chi cuadrado obtenido es de 0, por lo cual no rechazamos la hipótesis nula, o sea que los instrumentos son válidos.

## **Ejercicio 6**

Realizamos el test de Stock y Watson, que evalúa si el coeficiente del instrumento *nearc4* en la regresión de los residuos estimados en la primera etapa contra todas las variables exógenas es igual a cero. El estadístico F arrojó un valor de 17,61 y el p valor de 0,0000. Por lo tanto rechazamos la hipótesis nula de que el coeficiente es igual a cero, lo que aporta evidencia en contra de que el instrumento sea débil.

## **Ejercicio 7**

La principal crítica que puede realizarse al trabajo se centra en la validez del instrumento elegido. El autor argumenta que la distancia a las universidades no tiene efecto sobre el salario, sino sólo a través de su impacto en la educación. Esto implicaría entonces asumir que la distancia a la escuela no está correlacionada con el error, pero si con la educación. Sin embargo esta presunción no nos convence. Más allá de la estrategia implementada por el autor para testear la exogeneidad del instrumento, en verdad esta no puede testear empíricamente.

De hecho, al realizar el test de Hausman, no pudimos rechazar la hipótesis nula de que las varianzas de OLS y de IV sean iguales. Como sabemos que la educación es una variable endógena, la principal conclusión que podemos obtener de esto es que el supuesto de exogeneidad del instrumento no se cumple.

Precisamente podemos suponer que el hecho de que los padres que más se preocupan por la educación elijan vivir y criar a su familia cerca de los colegios y universidades, podría ser un mecanismo que una la distancia a la universidad y el salario causalmente. A su vez, también podría ser posible que los barrios más cercanos a la universidad sean más ricos y tengan más chances de acceder a salarios altos. En fin, no podemos aceptar así sin más la exogeneidad del instrumento.

**Código:**

*clear all*

*global main "C:\Users\juanb\OneDrive\Documentos\Juan\UdeSA\Economía Aplicada\Tutoriales\Tutorial 5"*

*global input "$main/input"*

*global output "$main/output"*

*cd "$main/input"*

*use "schooling-card.dta", clear*

*\*EJERCICIO 1*

*\*COLUMNA 2 de la TABLE 1 (los datos que están en la base son para esta submuestra)*

*\*Edad en 1966*

*gen interval = .*

*replace interval = 1 if age66==14 | age66==15*

*replace interval = 2 if age66==16 | age66==17*

*replace interval = 3 if age66>=18 & age66<=20*

*replace interval = 4 if age66>=21 & age66<=24*

*label define linterval 1 "14-15" 2 "16-17" 3 "18-20" 4 "21-24"*

*label value interval linterval*

*tab interval*

*\*REGIONES*

*gen region=.*

*replace south=1 if reg661==1 | reg662==1 | reg665==1*

*replace region=1 if reg661==1 | reg662==1 | reg665==1 /\*south\*/*

*replace region=2 if reg667==1 | reg669==1 /\*northeast\*/*

*replace region=3 if reg663==1 | reg664==1 /\*midwest\*/*

*replace region=4 if reg666==1 | reg668==1 /\*west\*/*

*label define lregion 1 "south" 2 "northeast" 3 "midwest" 4 "west"*

*label value region lregion*

*tab region*

*\*Vivió en SMSA en 1966*

*tab smsa76r*

*\*Vivió cerca de una Univ de 4 años en 1966*

*tab nearc4*

*\*Vivió con su madre y padre*

*tab momdad14*

*\*Vivió con su madre soltera*

*tab sinmom14*

*\*Educación promedio de la madre*

*sum momed*

*\*Educación promedio de la madre*

*sum daded*

*\*Porcentage de color negro*

*tab black*

*\*Promedio del score en el test KWW*

*sum kww*

*\*Educación promedio en 1976*

*sum ed76*

*\*Vivió en el sur en 1976:*

*sum reg76r*

*\*Submuestra 2: respuestas válidas de salario (empleados)*

*drop if wage76==.*

*tab smsa76r*

*tab nearc4*

*tab momdad14*

*tab sinmom14*

*tab black*

*tab region*

*sum momed daded ed76 reg76r*

*use "schooling-card.dta", clear*

*\*EJERCICIO 2*

*cd "$main/output"*

*\*Columna 1*

*reg lwage76 ed76 exp76 exp762 black south66 smsa76r, robust*

*est store mco1*

*outreg2 using MCO, word dec(4) label*

*\*Columna 2*

*reg lwage76 ed76 exp76 exp762 black south66 smsa66r reg66\* smsa66r, robust*

*est store mco2*

*outreg2 using MCO, word dec(4) label*

*\*Columna 3*

*reg lwage76 ed76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed, robust*

*est store mco3*

*outreg2 using MCO, word dec(4) label*

*\*Columna 4*

*reg lwage76 ed76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8, robust*

*est store mco4*

*outreg2 using MCO, word dec(4) label*

*\*Columna 5*

*reg lwage76 ed76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14, robust*

*est store mco5*

*outreg2 using MCO, word dec(4) label*

*\*EJERCICIO 3*

*\*COLUMNA 1*

*\*Basic speciication*

*quietly reg lwage76 ed76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14, robust*

*local beta1 = \_b[ed76]*

*display `beta1'*

*\*Use 1978 wages and education*

*quietly reg lwage78 ed76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14, robust*

*local beta2 = \_b[ed76]*

*display `beta2'*

*\*Include KWW Test Score*

*quietly reg lwage76 ed76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14 kww, robust*

*local beta3 = \_b[ed76]*

*display `beta3'*

*\*Use subsample age 14-19 in 1966*

*drop if age66<14 | age66>19*

*quietly reg lwage76 ed76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14, robust*

*local beta4 = \_b[ed76]*

*display `beta4'*

*cd "$main/input"*

*use "schooling-card.dta", clear*

*\*COLUMNA 2:*

*\*Basic speciication*

*quietly ivregress 2sls lwage76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14 (ed76= nearc4), robust*

*est store iv*

*local beta5 = \_b[ed76]*

*display `beta5'*

*\*Use 1978 wages and education*

*quietly ivregress 2sls lwage78 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14 (ed76= nearc4), robust*

*local beta6 = \_b[ed76]*

*display `beta6'*

*\*Include KWW Test Score*

*quietly ivregress 2sls lwage76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14 kww (ed76= nearc4), robust*

*local beta7 = \_b[ed76]*

*display `beta7'*

*\*Include KWW Test Score / Instrument with IQ*

*quietly ivregress 2sls lwage76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14 kww (ed76= nearc4 iq), robust*

*local beta8 = \_b[ed76]*

*display `beta8'*

*\*Use proximity to public college as isntrument for education*

*quietly ivregress 2sls lwage76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14 (ed76= nearc4a iq), robust*

*local beta9 = \_b[ed76]*

*display `beta9'*

*\*Use proximity to 2 year and 4 year college as isntrument for education*

*quietly ivregress 2sls lwage76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14 (ed76= nearc2 nearc4 iq), robust*

*local beta10 = \_b[ed76]*

*display `beta10'*

*\*Use subsample age 14-19 in 1966*

*drop if age66<14 | age66>19*

*quietly ivregress 2sls lwage76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14 (ed76= nearc4), robust*

*local beta11 = \_b[ed76]*

*display `beta11'*

*cd "$main/input"*

*use "schooling-card.dta", clear*

*\*EJERCICIO 4*

*hausman iv mco5, force*

*\*EJERCICIO 5*

*\* Sargan test*

*quietly ivregress 2sls lwage76 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14 (ed76 = nearc4), robust*

*predict resid, residual*

*reg resid nearc4 exp76 exp762 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14, robust*

*ereturn list*

*display chi2(2,e(N)\*e(r2))*

*\*EJERCICIO 6*

*\*first stage*

*reg ed76 nearc4 age66 black south66 smsa66r reg66\* smsa66r famed f1 f2 f3 f4 f5 f6 f7 f8 momdad14 sinmom14, robust*

*\*test*

*test nearc4==0*