

## Soluciones Ejercicios Calificados

### Tema 2

Para la realización de este ejercicio se requiere utilizar el software STATA, pero se permite el uso de otro programa estadístico siempre que pueda realizar lo solicitado. Se requiere la entrega de un informe en word o pdf (latex) con las respuestas a los ejercicios. Este informe podrá ser elaborado individualmente o por hasta 2 personas. Lo último para motivar la discusión.

### 1. Replicando a David Card, Premio Nobel de Economía 2021

Usted cuenta con la base de datos *card.dta*, correspondiente al paper “ Using Geographic Variation in College Proximity to Estimate the Return to Schooling” de David Card, 1993. Su primera tarea es leer el artículo, esto facilitará el desarrollo de esta tarea.

Sobre la base de los datos y la materia vista en clases, responda lo siguiente:

1. Realice una descripción de las características de los individuos de la muestra.

#### Respuesta

En la siguiente tabla se aprecian las características de los individuos que componen esta muestra:

<i>Distribución de la edad:</i>	
<i>Edad 14-15</i>	<i>25.5 %</i>
<i>Edad 16-17</i>	<i>24.1 %</i>
<i>Edad 18-20</i>	<i>24.6 %</i>
<i>Edad 21-24</i>	<i>25.8 %</i>
<i>Distribución Regional:</i>	
<i>Northeast</i>	<i>20.73 %</i>
<i>Midwest</i>	<i>26 %</i>
<i>South</i>	<i>41.43 %</i>
<i>West</i>	<i>11.9 %</i>
<i>Vivía en SMSA en 1966</i>	<i>65 %</i>
<i>Cercanía al Colegio</i>	<i>68.2 %</i>
<i>Estructura Familiar</i>	
<i>Madre y Padre</i>	<i>78.9 %</i>
<i>Sólo Madre</i>	<i>10.1 %</i>
<i>Promedio de la Educación de los Padres</i>	
<i>Educación de la Madre</i>	<i>10.3 años</i>
<i>Educación del Padre</i>	<i>10 años</i>
<i>Porcentaje de raza negra</i>	<i>23 %</i>
<i>Puntaje Promedio Test KWW</i>	<i>33.54</i>
<i>Media de Educación en 1976</i>	<i>13.3 años</i>
<i>Vivía en el sur en 1976</i>	<i>40.37 %</i>

2. Regresione mediante OLS, la variable lwage, respecto educ, exper, exper2, black, south, smsa, reg661-reg668, y smsa66. Comente los resultados obtenidos, y compárelos con la Tabla 2, Columna 2, del paper de Card (1993)

**Respuesta**

*La regresión de lwage respecto a las variables solicitadas se ilustra en el Cuadro 1. Al comparar los resultados con la Tabla 2 del paper de Card, vemos que esta especificación es la número (2) y los resultados son idénticos. El retorno a la educación es de 7.5% y vemos que es muy estable a especificaciones alternativas.*

Cuadro 1: Estimación OLS	
Variable	Coefficient (Std. Err.)
years of schooling, 1976	0.075** (0.004)
age - educ - 6	0.085** (0.007)
exper <sup>2</sup>	-0.002** (0.000)
=1 if black	-0.199** (0.018)
=1 if in south, 1976	-0.148** (0.028)
=1 in in SMSA, 1976	0.136** (0.019)
regional dummy, 1966	-0.119** (0.039)
reg662	-0.022 (0.030)
reg663	0.026 (0.028)
reg664	-0.063 <sup>†</sup> (0.037)
reg665	0.009 (0.039)
reg666	0.022 (0.041)
reg667	-0.001 (0.042)
reg668	-0.175** (0.047)
=1 if in SMSA, 1966	0.026 (0.019)
Intercept	4.739** (0.075)
N	3010
R <sup>2</sup>	0.3
F <sub>(15,2994)</sub>	91.31
Significance levels :    † : 10 %    * : 5 %    ** : 1 %	

3. El coeficiente de educación estimada que se presenta en la Tabla 2 de Card (1993), es bastante estable a través de las distintas especificaciones que se muestran, pero a pesar de esta estabilidad, podría suceder que esté sesgado el verdadero retorno económico por año de educación. Frente a esto, señale y explique al menos tres argumentos de las posibles fuentes de sesgo que se podrían ocasionar en este caso.

**Respuesta**

*En este contexto las posibles fuentes de sesgo son:*

- *Sesgo por habilidad: este sesgo se origina dado que carecemos de una medida de habilidad y esta queda en el término de error (sesgo de variable omitida). De acuerdo a la teoría del capital humano la correlación entre habilidad y educación es positiva y la correlación entre ingresos y habilidad también lo es. Esto implica que el sesgo por habilidad es positivo, luego el verdadero retorno a la educación debiese ser menor al 7.5 % encontrado.*
- *Error en la medida de educación: el sesgo por error de medida siempre es de atenuación (se puede demostrar).*
- *El verdadero retorno de la educación puede variar entre poblaciones. Si el retorno a la educación es heterogéneo, el retorno promedio calculado está influenciado por una subpoblación particular.*

4. Estime una ecuación para la variable educ de forma reducida, respecto a todas las variables explicatorias usadas en la parte b, incluyendo además la variable nearc4. ¿Qué podría decir respecto de la correlación parcial existente entre educ y nearc4?

**Respuesta**

*La estimación de la variable educ de forma reducida, se presenta en el Cuadro 2. Se puede ver que hay una correlación parcial importante entre educ y nearc4. De hecho si calculamos el estadístico  $t$  éste es de 3.7 (que al cuadrado nos da sobre 13 y equivale al estadístico de Cragg y Donald).*

Cuadro 2: Estimación Reducida de educ

Variable	Coefficient (Std. Err.)
age - educ - 6	-0.413** (0.032)
exper <sup>2</sup>	0.001 (0.002)
=1 if black	-0.936** (0.093)
=1 if in south, 1976	-0.052 (0.142)
=1 in in SMSA, 1976	0.402** (0.111)
regional dummy, 1966	0.303 (0.237)
reg662	0.224 (0.195)
reg663	0.275 (0.186)
reg664	0.420* (0.212)
reg665	0.030 (0.136)
reg666	0.000 (0.000)
reg667	0.086 (0.155)
reg668	0.827** (0.269)
reg669	0.513* (0.209)
=1 if in SMSA, 1966	0.025 (0.111)
=1 if near 4 yr college, 1966	0.320** (0.085)
Intercept	16.335** (0.225)
N	3010
R <sup>2</sup>	0.477
F (15,2994)	244.915
Significance levels : † : 10 % * : 5 % ** : 1 %	

5. Estime lwage por Variables Instrumentales, usando nearc4 como un instrumento para la variable educ. Comente sus resultados y compárelos con los obtenidos en 2.

### Respuesta

*En el Cuadro 3 podemos ver los resultados. El retorno a la educación aumenta a 13 %. Este valor es bastante alto y va en contra de la hipótesis del sesgo por habilidad puesto que si dicho sesgo fuese muy importante el retorno debiese haber bajado. Una explicación a esto son que la educación esté muy mal medida, luego el método nos ayuda a reducir dicho sesgo de atenuación (aunque se ha reportado en otros trabajos que el error de medida es por lo general bajo). La otra explicación*

*es el de hetoregeneidad del retorno a la educación y que el instrumento esté afectando a una subpoblación con un retorno particularmente alto (aquellos que no irían al college si no quedara cerca de la casa).*

Cuadro 3: Estimación por Variables Instrumentales

Variable	Coefficient (Std. Err.)
years of schooling, 1976	0.132 (0.055)
age - educ - 6	0.108 (0.024)
exper <sup>2</sup>	-0.002 (0.000)
=1 if black	-0.147 (0.054)
=1 if in south, 1976	-0.145 (0.027)
=1 in in SMSA, 1976	0.112 (0.032)
regional dummy, 1966	-0.108 (0.042)
reg662	-0.007 (0.033)
reg663	0.040 (0.032)
reg664	-0.058 (0.038)
reg665	0.038 (0.047)
reg666	0.055 (0.053)
reg667	0.027 (0.049)
reg668	-0.191 (0.051)
=1 if in SMSA, 1966	0.019 (0.022)
Intercept	3.774 (0.932)
N	3010
R <sup>2</sup>	0.238
F <sub>(15,.)</sub>	51.008
Significance levels : † : 10 % * : 5 % ** : 1 %	

6. Compare el intervalo de confianza a un 95 % del retorno de la educación del inciso 5 con el obtenido

en 2.

### Respuesta

*Intervalo de Confianza según regresión OLS:  $0,0675439 - 0,0818427$*

*Intervalo de Confianza según regresión VI:  $0,0240637 - 0,238944$*

*Claramente el método de VI es menos preciso y nos genera intervalos de confianza muy grandes. Con esto es muy difícil rechazar hipótesis nulas lo que nos puede llevar a cometer una mayor cantidad de errores de tipo II.*

7. En presencia de Instrumentos Débiles, el estimador por variables instrumentales está sesgado en la misma dirección que el estimador por OLS e incluso puede no ser consistente. Para saber si estamos o no ante la presencia de Instrumentos débiles, se le pide que testee esto usando el estadístico de Cragg y Donald (1993) y las tablas de Stock y Yogo (2005) con respecto a la medida del test de Wald .

### Respuesta

*En el siguiente Cuadro se reporta el estadístico de Cragg y Donald. Como mencionamos anteriormente este es mayor que 13 que parece ser un número razonable. Si bien el test de Stock y Yogo no nos garantiza que nuestra inferencia actual tenga a lo más 10 % de significancia en lugar 5 %. Si nos garantiza que a lo más tendremos una significancia de 15 %. Es probable que la significancia actual (empírica) esté más cerca del 10 % que del 15 % a juzgar por los valores de la tabla.*

<i>Cragg-Donald Wald F statistic):</i>		<i>13.256</i>
<i>Stock-Yogo weak ID test critical values:</i>	<i>10 % maximal IV size</i>	<i>16.38</i>
	<i>15 % maximal IV size</i>	<i>8.96</i>
	<i>20 % maximal IV size</i>	<i>6.66</i>
	<i>25 % maximal IV size</i>	<i>5.53</i>

8. Use `nearc2` con `nearc4` como instrumentos para `educ`. Primero estime la forma reducida para `educ`, y analice cuál de las dos variables está más fuertemente relacionada con `educ`. Después encuentre el estimador de VI que usa `nearc2` y `nearc4` como instrumento para `educ`. Discuta sus resultados. ¿Cambia su conclusión obtenida en la pregunta 7?

### Respuesta

*A continuación se presentan las estimaciones para la forma reducida y la estimación por 2SLS. En la forma reducida se ve claramente que `nearc4` es la única variable que está correlacionada con `educ`. La variable `nearc2` no está correlacionada con `educ`. La estimación por 2SLS muestra que el retorno de la educación es aún más alto que la estimación obtenida por VI con un único instrumento.*

Cuadro 4: Estimación reducida de educ incorporado nearc2 y nearc4

Variable	Coefficient (Std. Err.)
age - educ - 6	-0.412** (0.032)
exper <sup>2</sup>	0.001 (0.002)
=1 if black	-0.945** (0.093)
=1 if in south, 1976	-0.042 (0.142)
=1 in in SMSA, 1976	0.401** (0.111)
regional dummy, 1966	0.333 (0.237)
reg662	0.233 (0.195)
reg663	0.312 <sup>†</sup> (0.187)
reg664	0.465* (0.213)
reg665	0.065 (0.138)
reg666	0.000 (0.000)
reg667	0.125 (0.157)
reg668	0.884** (0.271)
reg669	0.502* (0.210)
=1 if in SMSA, 1966	0.000 (0.112)
=1 if near 4 yr college, 1966	0.321** (0.085)
=1 if near 2 yr college, 1966	0.123 (0.078)
Intercept	16.271** (0.229)
N	3010
R <sup>2</sup>	0.478
F <sub>(16,2993)</sub>	230.036
Significance levels :    † : 10 %    * : 5 %    ** : 1 %	



Cuadro 5: Estimación con nearc2 y nearc4 como instrumentos

Variable	Coefficient (Std. Err.)
years of schooling, 1976	0.157 (0.052)
age - educ - 6	0.119 (0.023)
exper <sup>2</sup>	-0.002 (0.000)
=1 if black	-0.123 (0.052)
=1 if in south, 1976	-0.143 (0.028)
=1 in in SMSA, 1976	0.101 (0.031)
regional dummy, 1966	-0.103 (0.043)
reg662	0.000 (0.034)
reg663	0.047 (0.033)
reg664	-0.055 (0.039)
reg665	0.052 (0.047)
reg666	0.070 (0.053)
reg667	0.039 (0.050)
reg668	-0.198 (0.052)
=1 if in SMSA, 1966	0.015 (0.022)
Intercept	3.340 (0.892)
N	3010
R <sup>2</sup>	0.17
F <sub>(15,.)</sub>	47.074
Significance levels : † : 10 % * : 5 % ** : 1 %	

9. ¿Cuándo las estimaciones por VI y MC2E son equivalentes? Analice si son equivalentes ambos

métodos según la muestra en estudio.

### Respuesta

*Las estimaciones por Variables Instrumentales y Mínimos Cuadrados en dos Etapas serán equivalentes cuando la cantidad de variables endógenas es la misma que el número de instrumentos, es decir cuando el modelo esté exactamente identificado.*

*Dado que en este modelo hay una variable endógena, entonces cuando sólo se utiliza un instrumento estos métodos son equivalentes.*

10. Sabemos que cuando hay más instrumentos que variables endógenas tenemos sobreidentificación, pues bien, considerando el caso del inciso 8, en el que se tienen dos instrumentos, realice el test de Sargan (1959) para restricciones de sobre-identificación.

(Ayuda: Para obtener el estadístico de Sargan, una opción es correr el modelo por MC2E, obteniendo  $\hat{u}$ . Luego corre la regresión auxiliar de  $\hat{u}$  sobre todas las variables exógenas ( $x$  y  $z$ 's) y obtiene el  $R^2$ . Se puede demostrar que  $\text{Sargan} = n \times R^2$ ).

### Respuesta

*A continuación se presentan los resultados del test de Sargan. Recuerde que la hipótesis nula es que los instrumentos son válidos y que el modelo está correctamente especificado. Podemos ver que el P-value es de 0.26 lo que implica que no se rechaza la hipótesis nula.*

*Tests of overidentifying restrictions:*

<i>Sargan <math>N \times R</math>-sq test</i>	<i>1.248 Chi-sq(1)</i>	<i>P-value = 0.2639</i>
<i>Basmann test</i>	<i>1.242 Chi-sq(1)</i>	<i>P-value = 0.2652</i>