

26

Ejercicios Calificados

Alumno: Giacomo Marinelli Tagliavento

Código: 20180462

Profesor: Tomás Rau Binder

Curso: Capacitación en Econometría Aplicada Avanzada 2022-01

Módulo: Variables Instrumentales

2022-01

Ejercicio 1:

El ejercicio 1 consiste en realizar una descripción de las características de los individuos de la muestra. Después de haber importado los datos en el stata, se ejecutó el comando “describe” para poder conocer las variables de las cuales disponemos. En este caso contamos con 3 010 observaciones. Se encontró que disponemos de varias de las variables del paper de David Card (1993), tales como las que describen la escolaridad de los padres, su inscripción en el colegio o no (considerando los diferentes años), su nivel de IQ, si se encuentra cerca a una universidad de acreditación de 4 o 2 años, edad, años de educación, entre otros. Es decir, poseemos un grupo de características individuales de las observaciones, un grupo de características del entorno familiar y un grupo de características que describen la ubicación geográfica del individuo.

El segundo paso fue realizar un summarize de toda la muestra en conjunto, resultados que podemos visualizar a continuación.

Variable	Obs	Mean	Std. Dev.	Min	Max
id	3,010	2581.749	1500.539	2	5225
nearc2	3,010	.4408638	.4965731	0	1
nearc4	3,010	.6820598	.4657535	0	1
educ	3,010	13.26346	2.676913	1	18
age	3,010	28.1196	3.137004	24	34
fatheduc	2,320	10.00345	3.720737	0	18
motheduc	2,657	10.34814	3.179671	0	18
weight	3,010	321185.3	170645.8	75607	1752340
momdad14	3,010	.7893688	.4078247	0	1
sinmom14	3,010	.1006645	.3009339	0	1
step14	3,010	.0388704	.1933182	0	1
south66	3,010	.4142857	.4926801	0	1
black	3,010	.2335548	.4231624	0	1
smsa	3,010	.7129568	.4524571	0	1
south	3,010	.4036545	.4907113	0	1
smsa66	3,010	.6495017	.4772053	0	1
wage	3,010	577.2824	262.9583	100	2404
enroll	3,010	.0923588	.2895799	0	1
kww	2,963	33.54067	8.611619	4	56
iq	2,061	102.4498	15.42376	50	149
married	3,003	2.271395	2.066823	1	6
libcrd14	2,997	.674341	.4686987	0	1
exper	3,010	8.856146	4.141672	0	23
lwage	3,010	6.261832	.4437976	4.60517	7.784889
expersq	3,010	95.57907	84.61831	0	529

Con ello se encontró que, en promedio, el 40% de las observaciones se encuentra cerca de una universidad con acreditación de 2 años y casi el 70% cerca de una universidad de acreditación de 4 años. El promedio de años de educación de las observaciones es de 13 años y, en promedio, poseen 28 años de edad (coincidiendo con el hecho de que las observaciones se encuentran entre el rango de 24 a 34 años). También se observó que los años de educación de sus padres es, en promedio, 10 años para ambos. En cuanto a la ubicación geográfica, encontramos que, en promedio, el 41% de las observaciones se encontró en la región Sur en el año 1966, pero solo el 40% en el año 1976; en cambio, el 65% se encontró en el SMSA ((metropolitan area) en el año 1966, pero casi el 71% se ubicaba en ese tipo de áreas en el año 1976. Por último, solo el 9% se inscribió en un colegio en el año 1976 (no hay una variable de inscripción para el año 1966).

En particular, contamos con la información casi toda la muestra válida, tanto para el salario como para la educación, del paper; sin embargo, no contamos con la información referente a la variable `midwest` que posee el paper, pero no resulta un problema en el análisis que realizaremos.

El tercer paso fue realizar otros comandos `summarize`, pero en este caso dividiendo las observaciones en 2 grupos: Primero una división entre las observaciones que vivieron en las regiones sur en el año 1966 o en el año 1976 y otro grupo de los individuos que vivieron en el SMSA en el año 1966 o en el 1976, segundo, una división entre las que vivieron cerca a solo a universidades con acreditación de 2 años y aquellos que vivieron cerca a solo universidades con acreditación de 4 años (en el año 1966).

Comparando en la primera división, observamos que las personas que vivieron en áreas metropolitanas estuvieron, en promedio, más cerca a universidades (ya sea acreditadas de 2 años o de 4). Asimismo, el 10% se logró inscribir a un colegio en 1976, en comparación al 9% (de las personas que vivieron en regiones del sur); además, el promedio de su salario (por hora) fue mayor en casi 1 dólar, en comparación a las personas de las regiones del sur. Por otra parte, ambos padres de los individuos que vivieron en el SMSA tuvieron un año más de educación y los mismo individuos también tuvieron un año más de educación. Los resultados los encontramos en los siguientes tablas.

Table 1: Individuos que vivieron en la región Sur en el 1966 o 1976

Variable	Obs	Mean	Std. Dev.	Min	Max
id	1,346	3587.895	1108.208	26	5225
nearc2	1,346	.3744428	.4841586	0	1
nearc4	1,346	.5594354	.4966394	0	1
educ	1,346	12.70728	2.852595	1	18
age	1,346	27.99703	3.060965	24	34
fatheduc	958	9.029228	4.252626	0	18
motheduc	1,148	9.570557	3.570352	0	18
weight	1,346	264778.9	202855	75607	1752340
momdad14	1,346	.7369985	.4404269	0	1
sinmom14	1,346	.1322437	.3388814	0	1
step14	1,346	.0319465	.1759231	0	1
south66	1,346	.9264487	.2611362	0	1
black	1,346	.4219911	.4940607	0	1
smsa	1,346	.641159	.4798387	0	1
south	1,346	.9026746	.2965105	0	1
smsa66	1,346	.5215453	.4997212	0	1
wage	1,346	506.0178	236.9378	100	2083
enroll	1,346	.0936107	.2913946	0	1
kww	1,323	31.01965	9.162324	4	56
iq	769	98.53186	16.53718	50	140
married	1,344	2.293899	2.067172	1	6
libcrd14	1,340	.5074627	.500131	0	1
exper	1,346	9.289747	4.313461	0	23
lwage	1,346	6.128954	.4415262	4.60517	7.641564
expersq	1,346	104.8915	92.82926	0	529

Table 2: Individuos que vivieron en SMSA en el 1966 o 1976

Variable	Obs	Mean	Std. Dev.	Min	Max
id	2,289	2325.287	1438.883	2	5219
nearc2	2,289	.491481	.5000367	0	1
nearc4	2,289	.7846221	.4111741	0	1
educ	2,289	13.60813	2.562973	1	18
age	2,289	28.13237	3.14154	24	34
fatheduc	1,791	10.43272	3.595904	0	18
motheduc	2,037	10.65292	3.027835	0	18
weight	2,289	335632.7	170213.8	75607	1752340
momdad14	2,289	.7964176	.4027498	0	1
sinmom14	2,289	.0978593	.2971893	0	1
step14	2,289	.0388816	.193355	0	1
south66	2,289	.3560507	.4789351	0	1
black	2,289	.2131935	.4096527	0	1
smsa	2,289	.9375273	.242065	0	1
south	2,289	.3425076	.4746521	0	1
smsa66	2,289	.8540848	.3530984	0	1
wage	2,289	612.5588	271.9974	100	2404
enroll	2,289	.1031018	.3041582	0	1
kww	2,250	34.59289	8.253247	6	56
iq	1,638	103.4451	15.40237	53	149
married	2,285	2.355799	2.112607	1	6
libcrd14	2,279	.7345327	.4416786	0	1
exper	2,289	8.524246	4.00409	0	23
lwage	2,289	6.325917	.4329946	4.60517	7.784889
expersq	2,289	88.68851	79.04694	0	529

Comparando en la segunda división encontramos lo siguiente. Las personas que vivieron, en el año 1966, cerca a solo universidades con acreditación de 2 años tuvieron un año de educación, en promedio, menor en comparación a los que estuvieron cerca a solo universidades con acreditación de 4 años, al igual que sus padres. También se inscribieron al colegio, en promedio, un 3% menos de personas y ganan 60 centavos de dólar menos por hora.

Table 3: Individuos cerca a solo universidades con acreditación de 2 años en 1966

Variable	Obs	Mean	Std. Dev.	Min	Max
id	339	3882.87	1136.055	949	5011
nearc2	339	1	0	1	1
nearc4	339	0	0	0	0
educ	339	12.32153	2.97522	2	18
age	339	27.81416	3.054385	24	34
fatheduc	238	8.878151	3.910101	0	18
motheduc	284	9.834507	3.317808	1	18
weight	339	266473.5	242639.3	75607	1752340
momdad14	339	.7522124	.432366	0	1
sinmom14	339	.1179941	.323078	0	1
step14	339	.0324484	.1774496	0	1
south66	339	.740413	.4390561	0	1
black	339	.4159292	.49361	0	1
smsa	339	.4572271	.4989035	0	1
south	339	.660767	.4741489	0	1
smsa66	339	.2625369	.4406632	0	1
wage	339	512.8555	246.3772	100	1763
enroll	339	.0648968	.2467077	0	1
kww	332	29.98494	9.347427	4	49
iq	190	98.97368	15.99386	60	133
married	339	2.339233	2.123403	1	6
libcrd14	339	.4631268	.4993756	0	1
exper	339	9.492625	4.502048	0	22
lwage	339	6.134062	.4637112	4.60517	7.474772
expersq	339	110.3186	98.13955	0	484

Table 4: Individuos cerca a solo universidades con acreditación de 4 años en 196

Variable	Obs	Mean	Std. Dev.	Min	Max
id	1,065	2383.469	1338.762	31	5225
nearc2	1,065	0	0	0	0
nearc4	1,065	1	0	1	1
educ	1,065	13.2939	2.625964	2	18
age	1,065	28.05728	3.118208	24	34
fatheduc	814	9.840295	3.656741	0	18
motheduc	941	10.26993	3.19698	0	18
weight	1,065	319127.4	162605.5	75607	1113393
momdad14	1,065	.7746479	.4180103	0	1
sinmom14	1,065	.1117371	.3151907	0	1
step14	1,065	.0366197	.1879145	0	1
south66	1,065	.4384977	.4964362	0	1
black	1,065	.2450704	.430331	0	1
smsa	1,065	.7380282	.439914	0	1
south	1,065	.4328638	.495705	0	1
smsa66	1,065	.6638498	.4726129	0	1
wage	1,065	568.7502	244.897	100	2083
enroll	1,065	.0929577	.2905096	0	1
kww	1,050	33.62286	8.382801	6	56
iq	712	102.4059	14.78204	54	139
married	1,064	2.162594	2.010057	1	6
libcrd14	1,059	.6761095	.4681799	0	1
exper	1,065	8.76338	4.127477	0	23
lwage	1,065	6.25567	.4242178	4.60517	7.641564
expersq	1,065	93.8169	84.56291	0	529

Ejercicio 2:

El segundo ejercicio implica realizar una regresión OLS con la variable dependiente lwage (el logaritmo del sueldo por hora en centavos) y las variables educ, exper, exper2, black, south, smsa, reg661-reg668 y smsa66 y compararlo con la Tabla 2 del paper de David Card (1993). De esta manera, los resultados obtenidos los podemos visualizar en el gráfico debajo. Nuestros resultados resultan ser bastante similares a los resultados obtenidos por el autor. Controlando por variables regionales y por la variable smsa66 (residencia en SMSA en 1966), el efecto estimado de la educación sobre los ingresos

(returns to schooling) es el de incrementar los ingresos en un aproximado de 7,5% por cada año adicional en educación, al igual que lo encontrado por los autores. También encontramos similitudes en el efecto de la experiencia, que es el incremento de un 8,5% de los salarios. Todas las demás variables, tales como black (indicador de la raza de la persona), south (región donde vive) y smsa, tienen los mismos efectos sobre el ingreso que lo observado por el autor con -20%, -15% y 14%, respectivamente (aproximando).

Estos resultados y los resultados obtenidos por el autor en sus columnas 1 y 3 (de la misma tabla, en donde agrega o elimina controles) son los que lo incitarán a utilizar variables instrumentales, debido a la posible endogeneidad de la variable educación; es decir, la existencia de una posible relación entre educación y factores no observados que influyen en los ingresos y en la educación al mismo tiempo (como la habilidad del individuo).

Source	SS	df	MS	Number of obs	=	3,010
				F(15, 2994)	=	85.48
Model	177.695591	15	11.8463727	Prob > F	=	0.0000
Residual	414.946054	2,994	.138592536	R-squared	=	0.2998
				Adj R-squared	=	0.2963
Total	592.641645	3,009	.196956346	Root MSE	=	.37228

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.0746933	.0034983	21.35	0.000	.0678339	.0815527
exper	.084832	.0066242	12.81	0.000	.0718435	.0978205
expersq	-.002287	.0003166	-7.22	0.000	-.0029079	-.0016662
black	-.1990123	.0182483	-10.91	0.000	-.2347927	-.1632318
south	-.147955	.0259799	-5.69	0.000	-.1988952	-.0970148
smsa	.1363845	.0201005	6.79	0.000	.0969724	.1757967
reg661	-.1185698	.0388301	-3.05	0.002	-.194706	-.0424335
reg662	-.0222026	.0282575	-0.79	0.432	-.0776088	.0332036
reg663	.0259703	.0273644	0.95	0.343	-.0276846	.0796251
reg664	-.0634942	.0356803	-1.78	0.075	-.1334546	.0064662
reg665	.0094551	.0361174	0.26	0.794	-.0613623	.0802725
reg666	.0219476	.0400984	0.55	0.584	-.0566755	.1005708
reg667	-.0005887	.0393793	-0.01	0.988	-.077802	.0766245
reg668	-.1750058	.0463394	-3.78	0.000	-.265866	-.0841456
smsa66	.0262417	.0194477	1.35	0.177	-.0118905	.0643739
_cons	4.739377	.0715282	66.26	0.000	4.599127	4.879626

Ejercicio 3:

El tercer ejercicio implica explicar al menos tres argumentos de las posibles fuentes de sesgo que se podrían ocasionar en la regresión que se realizó previamente (a través de OLS), específicamente en los resultados obtenidos del retorno económico por año de educación.

Una primera fuente de sesgo es la que también se detalló brevemente en el apartado anterior, en referencia a la posible relación entre la educación y las variables omitidas (o no observables) que también afectan a los ingresos, como lo es la variable de la habilidad de los individuos. Por ejemplo, un nivel de habilidad alto en los individuos puede ocasionar que estos reciban, en promedio, más ingresos a cualquier nivel de educación, por lo cual el estimador del retorno económico de los años de escolaridad estaría sesgado hacia arriba, pues la relación entre habilidad e ingresos es positiva y el coeficiente estaría absorbiendo este efecto. No considerar esta posible relación entre la variable educación y el componente no observado de la regresión ocasiona que esta variable sea endógena y su estimador no refleje el verdadero valor del efecto de un año más de educación en el ingreso.

Otra posible fuente de sesgo es el error de medición de la variable educación, lo cual implicaría que los valores que toma la variable en la regresión (los años de educación medidos) no son los reales valores de la variable (los años reales de educación del individuo). Normalmente el problema es el hecho de que el nivel de educación medido sea menor al verdadero valor de educación recibida por el individuo, es difícil que ocurra lo contrario, pues si ocurriese indicaría que los años de educación que el individuo dice poseer en realidad nunca los cumplió (algo ilógico dentro de una encuesta). Esto también ocasionaría que el efecto de la educación estimado no sea el verdadero, pues parte de los valores reales de educación se estarían incluyendo en el término no observado (en el término de error). Para explicar este hecho el autor cita a Griliches (1977) al ser uno de los primeros autores que mencionó esta problemática. Dicho autor planteó que, si se permite que la variable educación sea tratada de manera simétrica y tenga errores de medición, la relación entre el error y la educación puede hasta ser negativa, lo cual indicaría que el estimador OLS puede hasta estar subestimado.

Otra tercera fuente de sesgo es el hecho de que el retorno económico de la educación puede variar entre distintos grupos de población. Esto debido a que, si ciertos grupos de población reciben más retorno de la educación que otros grupos, entonces el efecto de los años de educación en el salario sin controlar por este tipo de características se estaría subestimando. Lo contrario ocurriría si ciertos grupos reciben menos retornos.

Sin embargo, David Card (1993) afirma que esta subestimación o sobreestimación dependerá del retorno económico de la educación de las familias con menores ingresos, debido a que se conoce que las familias con mayores ingresos (teniendo justamente más ingresos) tendrán retorno menor. De esta manera, si las familias con menores ingresos reciben un retorno por año adicional de educación mayor, la estimación por OLS estaría subestimando el efecto general del retorno por año adicional de educación.

Ejercicio 4:

El cuarto ejercicio implica estimar una ecuación para la variable educación de forma reducida con respecto a todas las variables explicativas usadas en el ejercicio 2, incluyendo además la variable `nearc4` (cuyo valor es 1 si el individuo vivió cerca a una universidad de acreditación de 4 años en el 1966), con la finalidad de poder dilucidar qué tipo de correlación parcial existe entre la educación y esta variable. Los resultados de dicha regresión los encontramos en el siguiente gráfico.

Source	SS	df	MS	Number of obs	=	3,010
Model	10287.6179	15	685.841194	F(15, 2994)	=	182.13
Residual	11274.4622	2,994	3.76568542	Prob > F	=	0.0000
				R-squared	=	0.4771
				Adj R-squared	=	0.4745
Total	21562.0801	3,009	7.16586243	Root MSE	=	1.9405

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
nearc4	.3198989	.0878638	3.64	0.000	.1476194	.4921785
exper	-.4125334	.0336996	-12.24	0.000	-.4786101	-.3464566
expersq	.0008686	.0016504	0.53	0.599	-.0023674	.0041046
black	-.9355287	.0937348	-9.98	0.000	-1.11932	-.7517377
south	-.0516126	.1354284	-0.38	0.703	-.3171548	.2139296
smsa	.4021825	.1048112	3.84	0.000	.1966732	.6076918
reg661	-.210271	.2024568	-1.04	0.299	-.6072395	.1866975
reg662	-.2889073	.1473395	-1.96	0.050	-.5778042	-.0000105
reg663	-.2382099	.1426357	-1.67	0.095	-.5178838	.0414639
reg664	-.093089	.1859827	-0.50	0.617	-.4577559	.2715779
reg665	-.4828875	.1881872	-2.57	0.010	-.8518767	-.1138982
reg666	-.5130857	.2096352	-2.45	0.014	-.9241293	-.1020421
reg667	-.4270887	.2056208	-2.08	0.038	-.8302611	-.0239163
reg668	.3136204	.2416739	1.30	0.194	-.1602434	.7874841
smsa66	.0254805	.1057692	0.24	0.810	-.1819071	.2328682
_cons	16.84852	.2111222	79.80	0.000	16.43456	17.26248

Según los resultados en el gráfico, podemos observar que, controlando por características individuales y geográficas, la variable *nearc4* resulta ser estadísticamente significativa para explicar los años de educación. Específicamente, vivir cerca a una universidad con acreditación de 4 años incrementa los años de educación en 0.32 (aproximadamente 4 meses más de educación). Este resultado es muy similar al que obtiene David Card (1993) en su tabla N°3 columna 1, a pesar de que nosotros solo hayamos agregado la variable *nearc4* y no en general (vivir cerca de cualquier universidad).

Sin embargo, el resultado sigue siendo útil para afirmar que el uso de *nearc4* como variable instrumental puede ser válida, ya que existe una relación estadísticamente significativa entre los años de educación y vivir cerca a este tipo de universidades. Faltaría pues, estimar la relación entre esta última variable y el componente no observado de la ecuación principal para asegurarnos la completa exogeneidad del instrumento y es justamente esta condición la que lleva al autor a utilizar diferentes controles y métodos para asegurarse de la robustez de su regresión por IV (Instrumental Variables).

Ejercicio 5:

El ejercicio 5 consiste en estimar el logaritmo del sueldo por hora a través del método de variables instrumentales, usando la variable `nearc4` como instrumento para la variable de años de educación. Los resultados los podemos observar en el siguiente gráfico.

Source	SS	df	MS	Number of obs	=	3,010
Model	141.146813	15	9.40978752	F(15, 2994)	=	51.01
Residual	451.494832	2,994	.150799877	Prob > F	=	0.0000
				R-squared	=	0.2382
				Adj R-squared	=	0.2343
Total	592.641645	3,009	.196956346	Root MSE	=	.38833

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.1315038	.0549637	2.39	0.017	.0237335	.2392742
exper	.1082711	.0236586	4.58	0.000	.0618824	.1546598
expersq	-.0023349	.0003335	-7.00	0.000	-.0029888	-.001681
black	-.1467757	.0538999	-2.72	0.007	-.2524603	-.0410912
south	-.1446715	.0272846	-5.30	0.000	-.19817	-.091173
smsa	.1118083	.031662	3.53	0.000	.0497269	.1738898
reg661	-.1078142	.0418137	-2.58	0.010	-.1898007	-.0258278
reg662	-.0070465	.0329073	-0.21	0.830	-.0715696	.0574767
reg663	.0404445	.0317806	1.27	0.203	-.0218694	.1027585
reg664	-.0579172	.0376059	-1.54	0.124	-.1316532	.0158189
reg665	.0384577	.0469387	0.82	0.413	-.0535777	.130493
reg666	.0550887	.0526597	1.05	0.296	-.0481642	.1583416
reg667	.026758	.0488287	0.55	0.584	-.0689832	.1224992
reg668	-.1908912	.0507113	-3.76	0.000	-.2903238	-.0914586
smsa66	.0185311	.0216086	0.86	0.391	-.0238381	.0609003
_cons	3.773965	.934947	4.04	0.000	1.940762	5.607169

Instrumented:	educ
Instruments:	exper expersq black south smsa reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668 smsa66 nearc4

Un resultado notorio de realizar esta nueva estimación a través de VI es el hecho de encontrar un estimador del coeficiente de educación más alto. A través de esta regresión, un año más de educación ocasiona un incremento de, aproximadamente, un 13% del salario por hora (en centavos) de los individuos. Esto también corresponde aproximadamente al efecto encontrado por los autores y dista mucho de ser igual a la

regresión por OLS realizada en el ejercicio 2. Teniendo en cuenta que en la regresión por OLS el retorno económico por un año adicional de educación era de un incremento del 7% en los salarios, podríamos decir que este estimador estaba subestimando el verdadero efecto de la variable. Las demás variables estadísticamente significativas son la experiencia, la experiencia al cuadrado, la raza y si la región donde se vive es sur o SMSA en 1976. Los efectos de estas variables en esta regresión por IV resultan ser levemente diferentes a lo encontrado con lo estimado en la regresión OLS, tal como el efecto de la variable south (-0.145 por IV y -0.147 por OLS), la variable smsa (0.1118 por IV y 0.136 por OLS) entre otros, lo cual sería otro indicador de que la variable educación sí presentaba un problema de endogeneidad.

Ejercicio 6:

El ejercicio 6 se basa en comparar los intervalos de confianza a un 95% de la estimación por OLS del ejercicio 2 y la estimación por IV del ejercicio 5. En la estimación de los años de educación a través de OLS encontramos un intervalo de confianza con los límites de 0.0678339 y 0.0815527; mientras que, en la estimación por IV se encontró un intervalo de confianza que tiene como límites 0.0237335 y 0.2392742. Ambos intervalos nos estarían indicando que en la estimación por IV la capacidad para encontrar significancia en el efecto de los años de escolaridad es menor, debido a que existe un rango muy grande de valores que podría, estadísticamente, tomar el coeficiente de educ. Esto también lo podemos observar comparando las desviaciones estándar de los modelos, en donde notamos que la estimación por OLS tiene una desviación estándar menor que la regresión por IV y que, en esta última, el p-value también ha crecido (pero igual resulta estadísticamente significativo). Todo esto nos podría dar un indicio de debilidad del instrumento.

Ejercicio 7:

El ejercicio 7 consiste en evaluar si, efectivamente, la variable nearc4 es un instrumento débil o no, a través del estadístico Cragg y Donald (1993) y las tablas de Stock y Yogo (2005) con respecto a la medida del test de Wald. Los resultados del tests los podemos

encontrar debajo de las estimaciones del siguiente gráfico.

Minimum eigenvalue statistic = 13.2558

Critical Values	# of endogenous regressors: 1			
Ho: Instruments are weak	# of excluded instruments: 1			
2SLS relative bias	5%	10%	20%	30%
	(not available)			
2SLS Size of nominal 5% Wald test	10%	15%	20%	25%
LIML Size of nominal 5% Wald test	16.38	8.96	6.66	5.53
	16.38	8.96	6.66	5.53

Según los resultados, tenemos un estadístico F correspondiente al estadístico de Cragg-Donald (el mínimo valor propio de la matriz de concentración) que, a su vez, corresponde al estadístico F de la primera etapa de los instrumentos (al tener una sola variable endógena). El valor de este estadístico es de 13.256. Los valores críticos correspondientes a una variable endógena y un instrumento exógeno son los que proporcionan Stock y Yogo y también se encuentran en el gráfico. En este caso, no podemos verificar los valores críticos que nos permitan verificar el grado de sesgo que estamos dispuestos a tolerar de la regresión de OLS, debido a que no nos encontramos en una situación de sobreidentificación. Por ello, debemos verificar la debilidad de los instrumentos a través del segundo test con valores críticos también elaborado por los mismos autores (Stock y Yogo). Este segundo test evalúa la significancia que estamos dispuestos a tolerar. Por ejemplo, en nuestro caso estamos dispuestos a tolerar un 15% de significancia empírica en vez de un 5% (el nivel de significancia teórico), pues nuestro minimo eigenvalue se encuentra entre los valores críticos de 16.38 (10% de significancia) y 8.96 (15% de significancia). Entonces nos encontramos en un nivel de inferencia aceptable, a pesar de no encontrarnos en un nivel de tolerancia del 10%. Con todo esto, podemos rechazar la hipótesis de instrumentos débiles, pero se podría mejorar la estimación encontrándonos en un nivel de tolerancia menor.

Ejercicio 8:

El ejercicio 8 consiste en agregar un nuevo instrumento a la variable educ. Para ello se solicita estimar la forma reducida para educ con los dos instrumentos y, después, encontrar el estimador de IV que use estas dos variables como instrumento para educ y nuevamente evaluar la debilidad de los instrumentos como en el ejercicio anterior. Los primeros resultados de la forma reducida para la variable educ se muestran a continuación.

Source	SS	df	MS	Number of obs	=	3,010
Model	10297.1164	16	643.569774	F(16, 2993)	=	170.99
Residual	11264.9637	2,993	3.76377002	Prob > F	=	0.0000
				R-squared	=	0.4776
				Adj R-squared	=	0.4748
Total	21562.0801	3,009	7.16586243	Root MSE	=	1.94

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
nearc2	.1229986	.0774256	1.59	0.112	-.0288142	.2748114
nearc4	.3205819	.0878425	3.65	0.000	.148344	.4928197
exper	-.4122915	.0336914	-12.24	0.000	-.4783521	-.3462309
expersq	.0008479	.00165	0.51	0.607	-.0023874	.0040832
black	-.9451729	.0939073	-10.06	0.000	-1.129302	-.7610434
south	-.0419115	.1355316	-0.31	0.757	-.3076561	.2238331
smsa	.4013708	.1047858	3.83	0.000	.1959113	.6068303
reg661	-.1687829	.2040832	-0.83	0.408	-.5689405	.2313747
reg662	-.269031	.1478324	-1.82	0.069	-.5588944	.0208325
reg663	-.1902114	.1457652	-1.30	0.192	-.4760216	.0955987
reg664	-.037715	.1891745	-0.20	0.842	-.4086403	.3332102
reg665	-.4371387	.1903306	-2.30	0.022	-.8103307	-.0639467
reg666	-.5022265	.2096933	-2.40	0.017	-.9133841	-.0910688
reg667	-.3775317	.207922	-1.82	0.070	-.7852162	.0301529
reg668	.3820043	.2454171	1.56	0.120	-.0991991	.8632076
smsa66	.0000782	.1069445	0.00	0.999	-.2096139	.2097704
_cons	16.77306	.2163481	77.53	0.000	16.34885	17.19727

Según la estimación de la forma reducida, pareciera ser que nearc2 (vivir cerca de una universidad con acreditación de 2 años) no resulta ser estadísticamente significativa para explicar los años de educación de los individuos. Sin embargo, aún se mantiene

la significancia estadística de la variable nearc4 (vivir cerca de una universidad con acreditación de 4 años) y su valor se vio modificado muy levemente. Esto coincide, en parte, a lo que mencionan los autores, ya que sus resultados demostraron el hecho de que, en la forma reducida de los años de educación, la variable nearc2 tiene un efecto muy pequeño en la educación. A continuación se muestra la estimación con el método de IV agregando estas dos variables como instrumentos para los años de educación.

```
Instrumental variables (2SLS) regression      Number of obs   =    3,010
                                             Wald chi2(15)   =    709.89
                                             Prob > chi2     =    0.0000
                                             R-squared       =    0.1702
                                             Root MSE       =    .4042
```

lwage	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
educ	.1570594	.0524383	3.00	0.003	.0542822	.2598366
exper	.1188149	.0227454	5.22	0.000	.0742348	.163395
expersq	-.0023565	.0003466	-6.80	0.000	-.0030358	-.0016772
black	-.1232778	.0520112	-2.37	0.018	-.225218	-.0213376
south	-.1431945	.0283691	-5.05	0.000	-.1987968	-.0875921
smsa	.100753	.0314355	3.21	0.001	.0391406	.1623654
reg661	-.102976	.0433068	-2.38	0.017	-.1878558	-.0180962
reg662	-.0002286	.0337043	-0.01	0.995	-.0662879	.0658306
reg663	.0469556	.0325621	1.44	0.149	-.016865	.1107763
reg664	-.0554084	.0390786	-1.42	0.156	-.132001	.0211842
reg665	.0515041	.0474412	1.09	0.278	-.0414789	.1444872
reg666	.0699968	.0531631	1.32	0.188	-.0342009	.1741945
reg667	.0390596	.0496175	0.79	0.431	-.0581889	.136308
reg668	-.1980371	.0523952	-3.78	0.000	-.3007297	-.0953444
smsa66	.0150626	.0222765	0.68	0.499	-.0285986	.0587238
_cons	3.339687	.8921571	3.74	0.000	1.591091	5.088283

Instrumented: educ

Instruments: exper expersq black south smsa reg661 reg662 reg663 reg664
reg665 reg666 reg667 reg668 smsa66 nearc4 nearc2

Observando los resultados de esta nueva estimación notamos que el coeficiente estimado para la variable educ es levemente mayor al obtenido en el ejercicio 5, cerca de un 2% mayor. Asimismo, notamos que la desviación estándar también se redujo levemente y, con ello, el valor del p-value es mucho menor al p-value con solo un instrumento

(al igual que se redujo, en cierta medida, su intervalo de confianza), lo cual le agrega consistencia a este estimador del retorno económico de un año más de educación. Estos resultados también son consistentes con lo obtenido por David Card, pues el hecho de incluir a `nearc2` como otra variable que podría aproximar la proximidad a una universidad no le otorgó una reducción de la desviación estándar tan grande. Ahora pues, es necesario evaluar si efectivamente las variables `nearc4` y `nearc2` son instrumentos débiles o no, a través del estadístico Cragg y Donald (1993) y las tablas de Stock y Yogo (2005). Los resultados se muestran a continuación.

Minimum eigenvalue statistic = 7.8931				
Critical Values	# of endogenous regressors: 1			
Ho: Instruments are weak	# of excluded instruments: 2			
	5%	10%	20%	30%
2SLS relative bias	(not available)			
2SLS Size of nominal 5% Wald test	10%	15%	20%	25%
LIML Size of nominal 5% Wald test	19.93	11.59	8.75	7.25
	8.68	5.33	4.42	3.92

Podemos observar que los resultados sobre la debilidad de los instrumentos nos otorga resultados no deseados. En particular nos seguimos encontrando en una situación que no nos permite ver el grado de sesgo de la estimación de OLS que estamos incluyendo en nuestra estimación, debido a que requeriríamos de un instrumento exógeno adicional. En cuanto a los valores críticos para evaluar la significancia, notamos que nuestro valor mínimo de la matriz de concentración es de 7.8931 que se encuentra entre los valores críticos de 8.75 y 7.25. Eso quiere decir que en nuestra estimación tolera hasta un 25% de significancia empírica en comparación al 5% de significancia teórica. Esto nos indicaría que podemos rechazar la hipótesis de instrumentos débiles; sin embargo, si comparamos este resultado con el obtenido en el ejercicio 7, entonces nos encontraríamos en una situación en donde los instrumentos nos proporcionan un nivel de significancia empírico que podría ser mucho mayor al teórico y a lo que tendríamos con solo un instrumento, lo cual es una situación más desfavorable (no sería recomendable incluir el segundo instrumento).

Ejercicio 9:

El ejercicio 9 implica explicar cuándo las estimaciones por VI y 2SLS son equivalentes y analizar si nuestra muestra de estudio nos permite tener estimaciones equivalente a través de ambos métodos.

Las estimaciones de por VI y 2SLS son equivalentes solo cuando el sistema está perfectamente identificado; es decir, cuando el número de instrumentos es igual al número de variables endógenas. Esto debido a que la fórmula de estimación a través de 2SLS permite que el vector de instrumentos (o el número de instrumentos) sea mayor al número de variables endógenas; en otras palabras, su fórmula de estimación es más general que la de VI, pues permite que el sistema sea sobreidentificado (número de instrumentos mayor al número de regresores endógenos), pero también que sea perfectamente identificado. Por otra parte, las estimaciones por VI solo nos otorga estimaciones consistentes si el sistema está perfectamente identificado.

De esta manera, en los ejercicios anteriores al 8avo, en los cuales tenemos un instrumento, deberíamos observar que las estimaciones del retorno económico por un año adicional de educación sean iguales. Para ello, se muestra a continuación los resultados, la figura 1 muestra los resultados a través de VI y la segunda a través de 2SLS.

Figure 1: Estimación por VI

Source	SS	df	MS	Number of obs	=	3,010
Model	141.146813	15	9.40978752	F(15, 2994)	=	51.01
Residual	451.494832	2,994	.150799877	Prob > F	=	0.0000
				R-squared	=	0.2382
				Adj R-squared	=	0.2343
Total	592.641645	3,009	.196956346	Root MSE	=	.38833

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.1315038	.0549637	2.39	0.017	.0237335	.2392742
exper	.1082711	.0236586	4.58	0.000	.0618824	.1546598
expersq	-.0023349	.0003335	-7.00	0.000	-.0029888	-.001681
black	-.1467757	.0538999	-2.72	0.007	-.2524603	-.0410912
south	-.1446715	.0272846	-5.30	0.000	-.19817	-.091173
smsa	.1118083	.031662	3.53	0.000	.0497269	.1738898
reg661	-.1078142	.0418137	-2.58	0.010	-.1898007	-.0258278
reg662	-.0070465	.0329073	-0.21	0.830	-.0715696	.0574767
reg663	.0404445	.0317806	1.27	0.203	-.0218694	.1027585
reg664	-.0579172	.0376059	-1.54	0.124	-.1316532	.0158189
reg665	.0384577	.0469387	0.82	0.413	-.0535777	.130493
reg666	.0550887	.0526597	1.05	0.296	-.0481642	.1583416
reg667	.026758	.0488287	0.55	0.584	-.0689832	.1224992
reg668	-.1908912	.0507113	-3.76	0.000	-.2903238	-.0914586
smsa66	.0185311	.0216086	0.86	0.391	-.0238381	.0609003
_cons	3.773965	.934947	4.04	0.000	1.940762	5.607169

Instrumented: educ

Instruments: exper expersq black south smsa reg661 reg662 reg663 reg664
reg665 reg666 reg667 reg668 smsa66 nearc4

Figure 2: Estimación por 2SLS

Instrumental variables (2SLS) regression	Number of obs	=	3,010
	Wald chi2(15)	=	769.20
	Prob > chi2	=	0.0000
	R-squared	=	0.2382
	Root MSE	=	.3873

lwage	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
educ	.1315038	.0548174	2.40	0.016	.0240637	.238944
exper	.1082711	.0235956	4.59	0.000	.0620246	.1545176
expersq	-.0023349	.0003326	-7.02	0.000	-.0029868	-.001683
black	-.1467757	.0537564	-2.73	0.006	-.2521364	-.0414151
south	-.1446715	.027212	-5.32	0.000	-.1980061	-.0913369
smsa	.1118083	.0315777	3.54	0.000	.0499171	.1736995
reg661	-.1078142	.0417024	-2.59	0.010	-.1895494	-.026079
reg662	-.0070465	.0328197	-0.21	0.830	-.0713719	.057279
reg663	.0404445	.031696	1.28	0.202	-.0216784	.1025675
reg664	-.0579172	.0375058	-1.54	0.123	-.1314272	.0155929
reg665	.0384577	.0468138	0.82	0.411	-.0532956	.130211
reg666	.0550887	.0525196	1.05	0.294	-.0478478	.1580252
reg667	.026758	.0486988	0.55	0.583	-.0686898	.1222058
reg668	-.1908912	.0505764	-3.77	0.000	-.2900191	-.0917634
smsa66	.0185311	.0215511	0.86	0.390	-.0237082	.0607704
_cons	3.773965	.9324588	4.05	0.000	1.946379	5.601551

Instrumented: educ

Instruments: exper expersq black south smsa reg661 reg662 reg663 reg664
reg665 reg666 reg667 reg668 smsa66 nearc4

Vemos que, efectivamente, los coeficientes estimados son iguales en ambas regresiones. La única diferencia que existe entre los parámetros estimados es su desviación estándar. De manera específica, las estimaciones hechas por 2SLS tienen una desviación estándar menor que aquellas hechas por OLS; sin embargo, las diferencias son casi insignificantes.

Ejercicio 10:

En este último ejercicio se utilizará nuevamente los dos instrumentos del ejercicio 8 (nearc2 y nearc4) esta vez para evaluar la restricción de sobre-identificación de Sargan

(el cual testea la validez de todos los instrumentos). Los resultados se muestran a continuación.

Tests of overidentifying restrictions:

Sargan (score) $\chi^2(1) = 1.24815$ (p = 0.2639)
Basmann $\chi^2(1) = 1.24162$ (p = 0.2652)

Según los resultados, no es posible rechazar la hipótesis nula de exogeneidad de los instrumentos, por lo que podemos afirmar que los instrumentos son exógenos.

En conclusión, nuestros resultados no distan de ser diferentes de los del paper de David Card (1993), tanto en la regresión de OLS como en las diferentes regresiones por IV, específicamente en esta última regresión en donde se utiliza una de sus especificaciones para medir la proximidad a una universidad con el fin de verificar la consistencia de su primer estimador con solo un instrumento (filas 5 y 6 de la tabla 4). Todos estos resultados coinciden con el hecho de que la regresión por OLS obtenía un efecto subestimado del retorno por año adicional de educación, debido al problema de endogeneidad de la variable años de educación con el término no observado (características no observadas de los individuos). Sin embargo, para verificar la validez de los instrumentos se han realizado pruebas estadísticas través del estadístico Cragg y Donald (1993) y las tablas de Stock y Yogo (2005) y el estadístico de Sargan (1959) en contraposición a las múltiples especificaciones de la regresión que ejecuta el autor, tales como el uso de características familiares o la interacción entre educación predicha (a través de probabilidades) y proximidad a la universidad como instrumentos.

Todos los resultados indicarían que la proximidad a una universidad con acreditación de 4 años sí es estadísticamente significativa para poder utilizarla como instrumento para exogeneizar los años de educación y evaluar el retorno económico de un año adicional de educación en la población, debido a que cumple con la condición de correlación con la variable endógena (instrumento fuerte) y de exogeneidad.

Bibliografia:

- David Card (1993). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. *National Bureau of Economic Research*. Working paper No. 4483.
- Zvi Griliches (1977). Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica*, volumen 45 (N°1), pp. 1-22.