

PRACTICO 2

**Omisión de variables, errores de medida
Variables Instrumentales**

Ejercicio 1 (Omisión de variables)

Parte I: Omisión de variables en el MRLS

Considere el siguiente modelo:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad \text{Modelo 1}$$

se cumple que $E(u|x_1, x_2) = 0$ (Supuesto 1)

se cuenta con una muestra aleatoria de observaciones iid $\{y_i, x_{1i}\}_{i=1,2,\dots,N}$ (Supuesto 2)

Se pide:

- 1) Escriba el modelo estimable (ponga especial atención a la forma del error de la ecuación).
- 2) Derive el sesgo asintótico que se produce cuando se estima β_1 a través de una regresión de y_i sobre x_{1i} ; ¿de qué depende dicho sesgo?
- 3) Analice en los siguientes casos el sesgo por omisión de variable (en el marco del Modelo 1, manteniendo los supuestos 1 y 2)
 - a. y : salario; x_1 : años de educación; x_2 : habilidad del individuo
 - b. y : salario; x_1 : años de educación; x_2 : edad
 - c. y : salario; x_1 : años de educación; x_2 : experiencia laboral
 - d. y : producción agrícola; x_1 : horas de trabajo; x_2 : calidad del suelo
 - e. y : producción agrícola; x_1 : horas de trabajo; x_2 : cantidad de heladas
 - f. y : consumo; x_1 : ingreso disponible; x_2 : riqueza
 - g. y : consumo; x_1 : ingreso disponible; x_2 : tasa subjetiva de descuento intertemporal

Razone:

- i. ¿es x_1 un regresor exógeno o endógeno?
- ii. ¿espera que haya sesgos en la estimación de los coeficientes?
- iii. en caso afirmativo: ¿es posible concluir respecto al signo de dicho sesgo?
- iv. ¿y respecto a la magnitud del sesgo?

Parte II: Omisión de variables en el MRLG

Considere el siguiente modelo:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i \quad \text{Modelo 2}$$

se cumple que $E(u|x_1, x_2, x_3) = 0$ (Supuesto 2.1)

se cuenta con una muestra aleatoria de observaciones iid $\{y_i, x_{1i}, x_{2i}\}_{i=1,2,\dots,N}$ (Supuesto 2.2)

Se pide:

- 1) Escriba el modelo estimable (ponga especial atención a la forma del error de la ecuación)
- 2) Indique el sesgo asintótico que se produce en la estimación de β_1 y β_2 a través de una regresión de y_i sobre x_{1i} y x_{2i} ¿de qué depende dicho sesgo?. En el caso que las variables x_2 y x_3 estén incorrelacionadas ¿se puede afirmar que la estimación de β_2 será consistente?
- 3) Analice en los siguientes casos el sesgo por omisión de variable (en el marco del Modelo 1, manteniendo los supuestos 1 y 2)
 - a. y : salario; x_1 : años de educación; x_2 : edad; x_3 : habilidad del individuo
 - b. y : salario; x_1 : años de educación; x_2 : experiencia laboral; x_3 : habilidad del individuo
 - c. y : producción agrícola; x_1 : horas de trabajo; x_2 : cantidad de heladas; x_3 : calidad del suelo

Razone:

- ii. ¿son x_1 y x_2 regresores exógenos o endógenos?
- iii. ¿espera que haya sesgos en la estimación de los coeficientes β_1 y β_2 ?

Ejercicio 2 (Errores de medida)

Considere el siguiente modelo:

$$y_i = \beta_0 + \beta_1 x_i^* + u_i \quad \text{Modelo 1}$$

se cumple que $E(u|x^*) = 0$ (Supuesto 1)

se cuenta con una muestra aleatoria de observaciones iid $\{y_i, x_i\}_{i=1,2,\dots,N}$ (Supuesto 2)

donde

$$x_i = x_i^* + \varepsilon_i \quad \text{se cumple que } \text{cov}(x^*, \varepsilon) = 0 \quad \text{cov}(u, \varepsilon) = 0 \quad (\text{Supuestos 3})$$

Se pide:

- 1) Escriba el modelo estimable (ponga especial atención a la forma del error de la ecuación).
- 2) Determine el sesgo asintótico que se produce cuando se estima β_1 a través de una regresión de y_i sobre x_i ¿de qué depende dicho sesgo?
- 3) Analice en los siguientes casos el sesgo por error de medida (en el marco del Modelo 1, manteniendo los supuestos 1 a 3)
 - a. y : consumo; x : ingreso disponible
 - b. y : salario; x : años de educación;
Razone:
 - a. ¿es x un regresor exógeno o endógeno?
 - b. ¿espera que haya un sesgo en la estimación de los coeficientes?
 - c. en caso afirmativo: ¿es posible concluir respecto al signo de dicho sesgo?
 - d. ¿y respecto a la magnitud del sesgo?

Ejercicio 3 (Regresores endógenos y estimación VI)

(Revisión Mayo 2012, Ejemplo 15.4 Wooldridge Introducción a la Econometría, 2ª edición)

En Card, D. (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," se propone la estimación por variables instrumentales de un modelo para estimar los retornos de la educación a partir de la siguiente ecuación:

$$\ln(w_i) = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{exper}_i + \beta_3 \text{exper}_i^2 + x_i' \gamma + u_i$$

Donde:

w : salario por hora

educ : años de educación completados

exper : experiencia potencial (edad-educ-6)

exper^2 : experiencia potencial al cuadrado sobre 100

El vector x incluye un conjunto de características observables tales como: Indicador de raza y dummies por lugar de residencia en el año 66 (son 9 regiones: reg661 a reg669).

Estimación 1

```
. reg lwage educ exper exper2 black reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668
```

Source	SS	df	MS	Number of obs =	3010
Model	158.799165	12	13.2332637	F(12, 2997) =	91.42
Residual	433.84248	2997	.144758919	Prob > F =	0.0000
				R-squared =	0.2680
				Adj R-squared =	0.2650
Total	592.641645	3009	.196956346	Root MSE =	.38047

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	.0790114	.0035504	22.25	0.000	.0720499 .0859729
exper	.0875037	.0067655	12.93	0.000	.0742381 .1007692
exper2	-.2440668	.0323233	-7.55	0.000	-.3074449 -.1806887
black	-.1828211	.018577	-9.84	0.000	-.219246 -.1463962
reg661	-.139752	.0396339	-3.53	0.000	-.2174643 -.0620397
reg662	-.0191121	.0288544	-0.66	0.508	-.0756886 .0374643
reg663	.0158491	.0279321	0.57	0.570	-.0389189 .0706172
reg664	-.1197227	.0358249	-3.34	0.001	-.1899666 -.0494789
reg665	-.1546054	.0286622	-5.39	0.000	-.2108049 -.0984059
reg666	-.1540476	.0332663	-4.63	0.000	-.2192746 -.0888205
reg667	-.1512786	.0317948	-4.76	0.000	-.2136206 -.0889367
reg668	-.198882	.0472951	-4.21	0.000	-.2916162 -.1061479
_cons	4.798294	.0725941	66.10	0.000	4.655954 4.940633

```
. est store MCO1
```

Estimación 2

```
. reg educ cercania exper exper2 black reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668
```

Source	SS	df	MS	Number of obs =	3010
Model	10197.5689	12	849.79741	F(12, 2997) =	224.10
Residual	11364.5112	2997	3.79196235	Prob > F =	0.0000
				R-squared =	0.4729
				Adj R-squared =	0.4708
Total	21562.0801	3009	7.16586243	Root MSE =	1.9473

educ	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
cercania	.4511111	.0804158	5.61	0.000	.2934354 .6087868
exper	-.4122875	.033806	-12.20	0.000	-.4785728 -.3460021
exper2	.0616767	.1654383	0.37	0.709	-.2627073 .3860608
black	-.917234	.0937926	-9.78	0.000	-1.101138 -.7333297
reg661	-.2626612	.2028006	-1.30	0.195	-.6603036 .1349813
reg662	-.2826386	.14774	-1.91	0.056	-.5723206 .0070435
reg663	-.2426988	.1429935	-1.70	0.090	-.5230741 .0376764
reg664	-.2062228	.1839477	-1.12	0.262	-.5668994 .1544538
reg665	-.5978626	.1473585	-4.06	0.000	-.8867966 -.3089287
reg666	-.6319188	.173194	-3.65	0.000	-.9715099 -.2923277
reg667	-.4815461	.1636825	-2.94	0.003	-.8024876 -.1606046
reg668	.2755212	.2423875	1.14	0.256	-.1997416 .750784
_cons	17.11109	.2037857	83.97	0.000	16.71151 17.51066

```
. est store MCO2
```

Estimación 3

```
. reg cercania educ exper exper2 black reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668
```

Source	SS	df	MS	Number of obs =	3010
Model	72.4409605	12	6.0367467	F(12, 2997) =	31.18
Residual	580.290269	2997	.193623713	Prob > F =	0.0000
				R-squared =	0.1110
				Adj R-squared =	0.1074
Total	652.731229	3009	.216926298	Root MSE =	.44003

cercania	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	.0230345	.0041062	5.61	0.000	.0149833 .0310856
exper	.0092538	.0078245	1.18	0.237	-.0060882 .0245958
exper2	-.0202356	.0373828	-0.54	0.588	-.0935342 .053063
black	.0760105	.0214848	3.54	0.000	.033884 .118137
reg661	.0208837	.0458377	0.46	0.649	-.0689929 .1107602
reg662	.0824481	.0333709	2.47	0.014	.0170158 .1478804
reg663	-.0670166	.0323043	-2.07	0.038	-.1303574 -.0036757
reg664	-.1883116	.0414325	-4.55	0.000	-.2695507 -.1070726
reg665	-.2192266	.0331486	-6.61	0.000	-.284223 -.1542302
reg666	-.4178311	.0384734	-10.86	0.000	-.4932681 -.3423942
reg667	-.2438311	.0367717	-6.63	0.000	-.3159313 -.1717309
reg668	-.1675116	.0546982	-3.06	0.002	-.2747614 -.0602619
_cons	.4244655	.0839573	5.06	0.000	.2598459 .5890852

```
. est store MCO3
```

Estimación 4

```
. ivreg lwage exper exper2 black reg661 reg662 reg663 reg664 reg665 reg666 reg667 reg668 (educ=cercania)
```

Instrumental variables (2SLS) regression

Source	SS	df	MS	Number of obs =	3010
Model	-86.6485253	12	-7.22071044	F(12, 2997) =	34.25
Residual	679.29017	2997	.226656713	Prob > F =	0.0000
				R-squared =	.
				Adj R-squared =	.
Total	592.641645	3009	.196956346	Root MSE =	.47608

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	.2252076	.0435823	5.17	0.000	.1397534 .3106618
exper	.1477947	.0197825	7.47	0.000	.1090059 .1865834
exper2	-.2518298	.0405116	-6.22	0.000	-.3312632 -.1723964
black	-.0523826	.0451294	-1.16	0.246	-.1408703 .0361052
reg661	-.1023405	.0508197	-2.01	0.044	-.2019855 -.0026955
reg662	.0171478	.0376728	0.46	0.649	-.0567193 .0910149
reg663	.0561695	.0369402	1.52	0.128	-.0162613 .1286003
reg664	-.0767075	.0466074	-1.65	0.100	-.1680932 .0146781
reg665	-.0516725	.0470966	-1.10	0.273	-.1440174 .0406725
reg666	-.0328478	.0549963	-0.60	0.550	-.1406821 .0749865
reg667	-.0638896	.0474812	-1.35	0.179	-.1569885 .0292094
reg668	-.2284216	.0598252	-3.82	0.000	-.3457243 -.111119
_cons	2.242164	.7634572	2.94	0.003	.7452112 3.739118

Instrumented: educ

Instruments: exper exper2 black reg661 reg662 reg663 reg664 reg665 reg666
reg667 reg668 cercania

```
. est store VI1
```

Contraste de Hausman

. hausman VI1 MC01

	—— Coefficients ——			
	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	VI1	MC01	Difference	S.E.
educ	.2252076	.0790114	.1461961	.0434374
exper	.1477947	.0875037	.060291	.0185897
exper2	-.2518298	-.2440668	-.007763	.0244212
black	-.0523826	-.1828211	.1304385	.0411286
reg661	-.1023405	-.139752	.0374115	.0318088
reg662	.0171478	-.0191121	.0362599	.0242211
reg663	.0561695	.0158491	.0403204	.0241739
reg664	-.0767075	-.1197227	.0430152	.0298132
reg665	-.0516725	-.1546054	.1029329	.0373707
reg666	-.0328478	-.1540476	.1211997	.0437944
reg667	-.0638896	-.1512786	.0873891	.035264
reg668	-.2284216	-.198882	-.0295396	.0366365

b = consistent under Ho and Ha; obtained from ivreg
 B = inconsistent under Ha, efficient under Ho; obtained from regress

Test: Ho: difference in coefficients not systematic

chi2(12) = (b-B)'[(V_b-V_B)^(-1)](b-B)
 = 11.33
 Prob>chi2 = 0.5011

Valor del estadístico del contraste = 11.33

p-valor de una χ^2 con 1 grado de libertad en 11.33 = .00076265
 p-valor de una χ^2 con 2 grados de libertad en 11.33 = .00346515
 p-valor de una χ^2 con 12 grados de libertad en 11.33 = .50086928
 p-valor de una χ^2 con 13 grados de libertad en 11.33 = .58319361

Se pide:

- 1) Defina "regresor endógeno" y señale qué fuentes de endogeneidad podrían estar afectando la variable *educ* en el modelo antes definido. Indique los potenciales sesgos y la dirección (signo) de los mismos.
- 2) Indique qué características debe tener una variable para ser un instrumento válido y que contrastes conoce para corroborar dicha validez.
- 3) En Card (1995) se propone utilizar la variable *cercanía* como instrumento para la variable *educ*, siendo la variable *cercanía* una dummy que indica si, al momento de tomar la decisión de realizar estudios universitarios (aprox. a los 18 años), los individuos residían en una localidad en la que había una institución terciaria que ofreciera carreras de 4 años. Discuta la validez de dicho instrumento, utilice para ello la evidencia que se ofrece en las estimaciones disponibles.

- 4) Realice un contraste para someter a prueba la hipótesis de que la variable *educ* es exógena. Indique la hipótesis nula y la alternativa del contraste, la forma del estadístico del contraste, su distribución asintótica y su conclusión. Analice las implicancias de dicha conclusión.
- 5) En base a la evidencia analizada señale cuál es la medida que le resulta más apropiada para medir cuánto se espera se incrementen (en promedio) los salarios ante un incremento de 1 año de educación.

Ejercicio 4 (Regresores endógenos y estimación VI)

(Examen diciembre 2009, Extraído parcialmente de Cameron y Trivedi , 2009, Microeconometrics using Stata)

Se desea investigar los determinantes del gasto en medicamentos en personas de más de 65 años en los Estados Unidos. Para ello se utiliza el Medical Expenditure Panel Survey (EUA) con información de personas de 65 años y más. El objetivo principal del estudio es determinar si el hecho de poseer un seguro que cubra los gastos en medicamentos promueve un mayor gasto en éstos.

La variable dependiente del modelo es:

GtoMedicam: Total de gastos en medicamentos (en logaritmos)

Los regresores a considerar son:

Seguro: Dummy que vale 1 si el individuo tiene un seguro de salud asociado a su empleador o al sindicato
Crónicas: Cantidad de enfermedades crónicas que tiene el individuo
Edad: Edad en años
Mujer: Dummy que vale 1 si la persona es de sexo femenino
Ingreso: Ingreso anual total del hogar (en logaritmos)

Además se cuenta con otras variables que no están incluidas en el modelo:

RatioSS: Ratio ingresos por la Seguridad Social (jubilaciones y pensiones) respecto al ingreso total.
IngresoBajo: Dummy que vale 1 si el ingreso del hogar está por debajo de determinado umbral.

Se realiza una estimación MCO pero se sospecha que ésta nos es una buena estimación ya que la variable Seguro podría ser endógena. El argumento es el siguiente: aunque la mayoría de los individuos en la muestra ya están retirados, es razonable suponer que aquellos que hubieran previsto mayores gastos en medicamentos habrían buscado un trabajo donde el empleador o el sindicato le proporcionaran un seguro para cubrir los gastos de medicamentos luego del retiro. Entonces, es razonable suponer que hay variables no observables que afectan tanto la variable *GtoMedicam* como la dummy *Seguro*. Así se procede a estimar el modelo utilizando el estimador de Variables Instrumentales, considerando dos instrumentos alternativos: *RatioSS* y *IngresoBajo*.

Se realizan las siguientes estimaciones:

Estimación 1: Estimación MCO de la variable GtosMedicam sobre los siguientes regresores: Seguro, Edad, Cronicas, Mujer e Ingreso .

```
. reg GtoMedicam Seguro Edad Cronicas Mujer Ingreso
```

Source	SS	df	MS	Number of obs =	10089
Model	3281.46741	5	656.293482	F(5, 10083) =	428.76
Residual	15433.6488	10083	1.5306604	Prob > F =	0.0000
				R-squared =	0.1753
				Adj R-squared =	0.1749
Total	18715.1162	10088	1.85518599	Root MSE =	1.2372

GtoMedicam	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Seguro	.0800203	.0260973	3.07	0.002	.0288644 .1311762
Edad	-.0027476	.0018797	-1.46	0.144	-.0064321 .000937
Cronicas	.4396048	.0095806	45.88	0.000	.420825 .4583847
Mujer	.05867	.0251863	2.33	0.020	.0092998 .1080402
Ingreso	.0195124	.0138181	1.41	0.158	-.0075738 .0465987
_cons	5.751526	.1513568	38.00	0.000	5.454836 6.048215

```
. est store MCO
```

Estimación 2: Estimación MCO variable dependiente Seguro sobre los siguientes regresores: RatioSS, Edad, Cronicas, Mujer e Ingreso.

```
reg Seguro RatioSS Edad Cronicas Mujer Ingreso
```

Source	SS	df	MS	Number of obs =	10089
Model	175.91949	5	35.1838981	F(5, 10083) =	160.79
Residual	2206.32335	10083	.218816161	Prob > F =	0.0000
				R-squared =	0.0738
				Adj R-squared =	0.0734
Total	2382.24284	10088	.236146197	Root MSE =	.46778

Seguro	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
RatioSS	-.1937974	.0141383	-13.71	0.000	-.2215112 -.1660836
Edad	-.0083182	.0007109	-11.70	0.000	-.0097118 -.0069246
Cronicas	.0125198	.0036263	3.45	0.001	.0054116 .019628
Mujer	-.0732171	.009504	-7.70	0.000	-.0918468 -.0545874
Ingreso	.0520009	.0056354	9.23	0.000	.0409543 .0630474
_cons	.986775	.0568269	17.36	0.000	.8753829 1.098167

Estimación 3: Estimación MCO variable dependiente Seguro sobre los siguientes regresores: IngresoBajo, Edad, Cronicas, Mujer e Ingreso.

```
. reg Seguro IngresoBajo Edad Cronicas Mujer Ingreso
```

Source	SS	df	MS	Number of obs =	10089
Model	147.043485	5	29.4086969	F(5, 10083) =	132.66
Residual	2235.19935	10083	.221679991	Prob > F =	0.0000
				R-squared =	0.0617
				Adj R-squared =	0.0613
Total	2382.24284	10088	.236146197	Root MSE =	.47083

Seguro	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
IngresoBajo	-.0915806	.012326	-7.43	0.000	-.115742 -.0674192
Edad	-.0093884	.0007097	-13.23	0.000	-.0107797 -.0079972
Cronicas	.0102067	.003645	2.80	0.005	.0030618 .0173516
Mujer	-.0789788	.0095534	-8.27	0.000	-.0977054 -.0602522
Ingreso	.0748563	.0053096	14.10	0.000	.0644485 .0852641
_cons	.9252111	.0569487	16.25	0.000	.8135803 1.036842

Estimación 4: Estimación VI de la variable GtoMedicam sobre los siguientes regresores: Seguro, Edad, Cronicas, Mujer e Ingreso utilizando RatioSS como instrumento de la variable “Seguro”

```
. ivreg GtoMedicam Edad Cronicas Mujer Ingreso ( Seguro = RatioSS)
```

Instrumental variables (2SLS) regression

Source	SS	df	MS	Number of obs =	10089
Model	1299.63019	5	259.926039	F(5, 10083) =	381.82
Residual	17415.486	10083	1.72721274	Prob > F =	0.0000
				R-squared =	0.0694
				Adj R-squared =	0.0690
Total	18715.1162	10088	1.85518599	Root MSE =	1.3142

GtoMedicam	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Seguro	-.859033	.2049656	-4.19	0.000	-1.260806 -.4572596
Edad	-.0118072	.0027974	-4.22	0.000	-.0172907 -.0063236
Cronicas	.4488576	.010372	43.28	0.000	.4285264 .4691888
Mujer	-.016775	.0313372	-0.54	0.592	-.078202 .0446521
Ingreso	.0975094	.0223603	4.36	0.000	.0536788 .14134
_cons	6.602719	.2444106	27.01	0.000	6.123626 7.081813

Instrumented: Seguro

Instruments: Edad Cronicas Mujer Ingreso RatioSS

```
. est store VI_RATIOSS
```

Estimación 5: Estimación VI de la variable GtoMedicam sobre los siguientes regresores: Seguro, Edad, Cronicas, Mujer e Ingreso utilizando IngresoBajo como instrumento de la variable “Seguro”

```
. ivreg GtoMedicam Edad Cronicas Mujer Ingreso ( Seguro = IngresoBajo)
```

Instrumental variables (2SLS) regression

Source	SS	df	MS	Number of obs =	10089
Model	3274.09787	5	654.819574	F(5, 10083) =	426.71
Residual	15441.0184	10083	1.53139129	Prob > F =	0.0000
				R-squared =	0.1749
				Adj R-squared =	0.1745
Total	18715.1162	10088	1.85518599	Root MSE =	1.2375

GtoMedicam	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Seguro	.1372836	.3537511	0.39	0.698	-.556139 .8307062
Edad	-.0021951	.0038883	-0.56	0.572	-.009817 .0054268
Cronicas	.4390406	.0101939	43.07	0.000	.4190586 .4590226
Mujer	.0632706	.037921	1.67	0.095	-.0110621 .1376033
Ingreso	.0147562	.0323983	0.46	0.649	-.0487508 .0782632
_cons	5.69962	.3538058	16.11	0.000	5.00609 6.39315

Instrumented: Seguro

Instruments: Edad Cronicas Mujer Ingreso IngresoBajo

```
. est store VI_IngresoBajo
```

Contraste 1: Se realiza un contraste de Hausman utilizando la estimación 4 como estimación consistente y se la compara con la regresión MCO de la estimación 1

```
hausman VI_RATIOSS MCO
```

	---- Coefficients ----			
	(b) RATIO	(B) .	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
Seguro	-.859033	.0800203	-.9390533	.2032974
Edad	-.0118072	-.0027476	-.0090596	.0020718
Cronicas	.4488576	.4396048	.0092528	.0039737
Mujer	-.016775	.05867	-.0754449	.0186458
Ingreso	.0975094	.0195124	.0779969	.0175796

b = consistent under Ho and Ha; obtained from ivreg
B = inconsistent under Ha, efficient under Ho; obtained from regress

Test: Ho: difference in coefficients not systematic

```
chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 21.34
Prob>chi2 = 0.0007
```

Contraste 2: Se realiza un contraste de Hausman utilizando la estimación 5 como estimación consistente y se la compara con la regresión MCO de la estimación 1

. hausman VI_IngresoBajo MCO

	---- Coefficients ----		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) IB	(B) .		
Seguro	.1372836	.0800203	.0572633	.3527871
Edad	-.0021951	-.0027476	.0005525	.0034038
Cronicas	.4390406	.4396048	-.0005642	.0034824
Mujer	.0632706	.05867	.0046006	.0283487
Ingreso	.0147562	.0195124	-.0047562	.0293037

b = consistent under Ho and Ha; obtained from ivreg
 B = inconsistent under Ha, efficient under Ho; obtained from regress

Test: Ho: difference in coefficients not systematic

chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
 = 0.03
 Prob>chi2 = 1.0000

Se Pide:

- 1) Interprete los resultados de la estimación 1 (recuerde cual es el objetivo principal del estudio a la hora de dedicar tiempo a la interpretación de las distintas variables). Señale cuál sería el problema del estimador MCO para estimar el efecto de la variable Seguro en los gastos en medicamentos.
- 2) Señale que características debe tener una variable para ser un instrumento válido. Analice y discuta la validez de las variables RatioSS y IngresoBajo como instrumentos para la variable Seguro en el modelo a estimar.
- 3) Compare las estimaciones del coeficiente asociado a la variable Seguro en las estimaciones 1 y 4 por un lado y 1 y 5 por el otro. Interprete los resultados.
- 4) Complemente sus comentarios del punto 4 indicando que conclusiones extrae de los contrastes de Hausman realizados. En particular, señale si la evidencia analizada permite concluir que la variable Seguro es endógena.
- 5) ¿Le permite la evidencia analizada concluir que poseer un seguro que cubra los gastos en medicamentos promueve un mayor gasto en éstos? Justifique

Ejercicio 5 (Regresores endógenos y estimación VI)
 (Examen Mayo 2008)

Considere un modelo en el cual el gasto en salud está dado por la suma de una proporción constante del ingreso permanente y un componente aleatorio que suponemos no correlacionado con el ingreso permanente:

$$gs_i = \beta y_i^P + v_i \quad (1)$$

Se observa el gasto en salud y el ingreso disponible definido como

$$y_i = y_i^P + y_i^T \quad (2)$$

donde el supraíndice P denota el componente permanente y el T el componente transitorio del ingreso. Los subíndices i corresponden a los diferentes individuos en la muestra.

La Encuesta de Ingresos y Gastos realizada por el Instituto de Estadísticas y Censos del Uruguay proporciona datos respecto al ingreso disponible y los gastos de salud para un conjunto de 1916 hogares.

Utilizando esa muestra se procedió a realizar una regresión del gasto en salud (mlsal) sobre el ingreso disponible (mling), ambas variables en logaritmos y considerando los desvíos respecto a las respectivas medias muestrales.

Los resultados fueron:

Linear regression	Number of obs =	1916
	F(1, 1914) =	400.20
	Prob > F =	0.0000
	R-squared =	0.2206
	Root MSE =	1.997

mlsal	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
mling	1.474354	.0736997	20.00	0.000	1.329813	1.618894
_cons	2.47e-07	.0456227	0.00	1.000	-.0894752	.0894757

Dada las características del modelo antes detallado se sospecha que en dicha estimación el regresor “ingreso disponible” podría ser endógeno. Para ello se propone recurrir a una estimación por variables instrumentales utilizando como instrumentos el gasto en vivienda y el gasto en enseñanza, ambas variables también en logaritmos y considerando los desvíos respecto a las respectivas medias muestrales.

Para proceder a la estimación en primer lugar se procede a realizar una regresión MCO del ingreso disponible sobre el gasto en vivienda y el gasto en enseñanza. Los resultados son:

Source	SS	df	MS	Number of obs =	1916
Model	194.034674	2	97.017337	F(2, 1913) =	232.07
Residual	799.745304	1913	.418058183	Prob > F =	0.0000
Total	993.779977	1915	.518945158	R-squared =	0.1952
				Adj R-squared =	0.1944
				Root MSE =	.64657

mling	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
mlviv	.0724792	.0063199	11.47	0.000	.0600845	.0848739
mlens	.1077466	.0062227	17.32	0.000	.0955426	.1199506
_cons	-1.14e-08	.0147714	-0.00	1.000	-.0289697	.0289697

A continuación se estima un modelo de variables instrumentales en el que el gasto en vivienda y el gasto en enseñanza se utilizan como instrumentos del gasto en salud. Los resultados son:

Instrumental variables (2SLS) regression

```
-----
Total (centered) SS      = 9793.281472
Total (uncentered) SS   = 9793.281472
Residual SS             = 7758.868727

Number of obs = 1916
F( 1, 1914) = 160.32
Prob > F      = 0.0000
Centered R2   = 0.2077
Uncentered R2 = 0.2077
Root MSE      = 2.012
```

mlsal	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
miling	1.830124	.1444647	12.67	0.000	1.546978	2.113269
_cons	2.53e-07	.0459731	0.00	1.000	-.0901053	.0901059

```
Instrumented:      miling
Instruments:      mlviv mlens
-----
```

Finalmente se realiza un contraste de Hausman para probar la endogeneidad del regresor "ingreso disponible", los resultados son:

	---- Coefficients ----			
	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	.	MCO	Difference	S.E.
miling	1.830124	1.474354	.3557702	.1242514

```
-----
b = consistent under Ho and Ha; obtained from ivreg2
B = inconsistent under Ha, efficient under Ho; obtained from regress
```

```
chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
        = 8.20
Prob>chi2 = 0.0042
```

Se pide:

- 1) Derive el potencial sesgo de la estimación del parámetro β utilizando una regresión MCO del gasto en salud sobre el ingreso disponible.
- 2) Discuta la validez de utilizar el gasto en vivienda y el gasto en salud como instrumentos para el ingreso disponible.
- 3) Describa el procedimiento de estimación por Mínimos Cuadrados en dos etapas y determine la forma del estimador. Señale condiciones necesarias para que el modelo esté identificado.
- 4) Explique en que consiste el contraste de Hausman detallando la hipótesis nula y la alternativa del contraste, el estadístico del contraste y su distribución asintótica. Aplique dicho contraste utilizando las estimaciones realizadas.

Ejercicio 6

Ejercicio 5.14 en Wooldridge, J. M. (2006) *Introducción a la Econometría: un enfoque moderno*, Thomson Learning, México, 2ª. Edición, página 580.

Ejercicio 7

Problema 5.3 en Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT Press, página 116.