

Trabajo Práctico N° 6:

Modelos para Variables Dependientes Limitadas - Heckman.

Ejercicio 1: Gastos Ambulatorios.

Retomar la base de datos del Ejercicio 2 del Problem Set 5. Ahora, se estimará un modelo de dos partes de Heckman. Estos modelos sirven para muestras autoseleccionadas. Se modela, explícitamente, la ecuación que determina la selección y la ecuación de interés. En este ejercicio, se pide estimar un modelo de Heckman para los gastos ambulatorios y comparar con las predicciones de un modelo Tobit.

Heckman (MLE):

Heckman selection model	Number of obs	=	3,328
(regression model with sample selection)	Selected	=	2,802
	Nonselected	=	526

Log likelihood = -5836.219	Wald chi2(6)	=	288.88
	Prob > chi2	=	0.0000

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	

lambexp						
age	.2119749	.0230072	9.21	0.000	.1668816	.2570682
female	.3481441	.0601142	5.79	0.000	.2303223	.4659658
educ	.018716	.0105473	1.77	0.076	-.0019563	.0393883
blhisp	-.2185714	.0596687	-3.66	0.000	-.3355199	-.101623
totchr	.53992	.0393324	13.73	0.000	.4628299	.61701
ins	-.0299871	.0510882	-0.59	0.557	-.1301182	.0701439
_cons	5.044056	.2281259	22.11	0.000	4.596938	5.491175

dambexp						
age	.0879359	.027421	3.21	0.001	.0341917	.14168
female	.6626649	.0609384	10.87	0.000	.5432278	.7821021
educ	.0619485	.0120295	5.15	0.000	.0383711	.0855258
blhisp	-.3639377	.0618734	-5.88	0.000	-.4852073	-.2426682
totchr	.7969518	.0711306	11.20	0.000	.6575383	.9363653
ins	.1701367	.0628711	2.71	0.007	.0469117	.2933618
income	.0027078	.0013168	2.06	0.040	.000127	.0052886
_cons	-.6760546	.1940288	-3.48	0.000	-1.056344	-.2957652

/athrho	-.1313456	.1496292	-0.88	0.380	-.4246134	.1619222
/lnsigma	.2398173	.0144598	16.59	0.000	.2114767	.268158

rho	-.1305955	.1470772			-.4008098	.1605217
sigma	1.271017	.0183786			1.235501	1.307554
lambda	-.1659891	.1878698			-.5342072	.2022291

LR test of indep. eqns. (rho = 0): chi2(1) = 0.91				Prob > chi2 = 0.3406		

```
Heckman selection model -- two-step estimates      Number of obs   =       3,328  
(regression model with sample selection)           Selected        =       2,802  
                                                    Nonselected     =         526  
  
Wald chi2(6)          =       193.43  
Prob > chi2            =       0.0000
```

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
<hr/>						
lambexp						
age	.2024668	.0242202	8.36	0.000	.1549961	.2499374
female	.2921341	.0725756	4.03	0.000	.1498886	.4343796
educ	.0123889	.0115682	1.07	0.284	-.0102844	.0350622
blhisp	-.1828659	.0653449	-2.80	0.005	-.3109396	-.0547922
totchr	.5006332	.0485548	10.31	0.000	.4054675	.5957988
ins	-.0465097	.0529742	-0.88	0.380	-.1503373	.0573179
_cons	5.288927	.288522	18.33	0.000	4.723435	5.85442
<hr/>						
dambexp						
age	.0868152	.0274556	3.16	0.002	.0330032	.1406272
female	.6635053	.0609648	10.88	0.000	.5440165	.7829941
educ	.061884	.012039	5.14	0.000	.038288	.0854801
blhisp	-.3657835	.0619095	-5.91	0.000	-.4871239	-.2444432
totchr	.7957496	.0712174	11.17	0.000	.656166	.9353332
ins	.169107	.0629296	2.69	0.007	.0457673	.2924467
income	.0026773	.0013105	2.04	0.041	.0001088	.0052458
_cons	-.6686471	.1941247	-3.44	0.001	-1.049125	-.2881698
<hr/>						
/mills						
lambda	-.4637133	.2825997	-1.64	0.101	-1.017598	.090172
<hr/>						
rho	-0.35907					
sigma	1.2914258					

Tobit regression	Number of obs	=	2,802
	Uncensored	=	2,801
Limits: Lower = 0	Left-censored	=	1
Upper = +inf	Right-censored	=	0
	LR chi2(6)	=	596.53
	Prob > chi2	=	0.0000
Log likelihood = -4642.5217	Pseudo R2	=	0.0604

lambexp	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
age	.2172778	.0222024	9.79	0.000	.1737431	.2608126
female	.3795502	.0485335	7.82	0.000	.2843851	.4747153
educ	.0221958	.0097527	2.28	0.023	.0030726	.0413191
blhisp	-.2384675	.0551452	-4.32	0.000	-.346597	-.1303381
totchr	.5618619	.0304802	18.43	0.000	.502096	.6216278
ins	-.0210413	.0499613	-0.42	0.674	-.119006	.0769234
_cons	4.908076	.1679989	29.21	0.000	4.578661	5.23749
var(e.lambexp)	1.608909	.0429988			1.526767	1.69547

Tabla comparativa:

	(1) Heckman (M~)	(2) Heckman (T~)	(3) Tobit
lambexp			
age	0.212*** (0.0230)	0.202*** (0.0242)	0.217*** (0.0222)
female	0.348*** (0.0601)	0.292*** (0.0726)	0.380*** (0.0485)
educ	0.0187* (0.0105)	0.0124 (0.0116)	0.0222** (0.00975)
blhisp	-0.219*** (0.0597)	-0.183*** (0.0653)	-0.238*** (0.0551)
totchr	0.540*** (0.0393)	0.501*** (0.0486)	0.562*** (0.0305)
ins	-0.0300 (0.0511)	-0.0465 (0.0530)	-0.0210 (0.0500)
_cons	5.044*** (0.228)	5.289*** (0.289)	4.908*** (0.168)
dambexp			
age	0.0879*** (0.0274)	0.0868*** (0.0275)	
female	0.663*** (0.0609)	0.664*** (0.0610)	
educ	0.0619*** (0.0120)	0.0619*** (0.0120)	
blhisp	-0.364*** (0.0619)	-0.366*** (0.0619)	
totchr	0.797*** (0.0711)	0.796*** (0.0712)	
ins	0.170*** (0.0629)	0.169*** (0.0629)	
income	0.00271** (0.00132)	0.00268** (0.00131)	
_cons	-0.676*** (0.194)	-0.669*** (0.194)	
/			
athrho	-0.131 (0.150)		
lnsigma	0.240*** (0.0145)		
var(e.lamb~)			1.609*** (0.0430)
/mills			
lambda		-0.464 (0.283)	
N	3328	3328	2802
pseudo R-sq			0.060

Standard errors in parentheses
 * p<0.10, ** p<0.05, *** p<0.01

Ejercicio 2: Ecuación Salarial para las Mujeres I.

Considerar la base de datos “womenwk.dta”. Describir la base. Estimar una ecuación salarial en función de la educación y la edad por Mínimos Cuadrados Clásicos. Repetir utilizando un modelo de Heckman, utilizando las variables married, children, education y age para la ecuación de selección. Utilizar el comando heckman.

Descripción de la base:

Variable	Obs	Mean	Std. dev.	Min	Max
county	2,000	4.5	2.873	0	9
age	2,000	36.208	8.28656	20	59
education	2,000	13.084	3.045912	10	20
married	2,000	.6705	.4701492	0	1
children	2,000	1.6445	1.398963	0	5
wage	1,343	23.69217	6.305374	5.88497	45.80979

Hourly wage; missing, if not working

Percentiles		Smallest		
1%	9.728734	5.88497		
5%	13.48302	6.739784		
10%	15.69925	7.12612	Obs	1,343
25%	19.30873	7.328383	Sum of wgt.	1,343
			Mean	23.69217
50%	23.51122		Std. dev.	6.305374
		Largest		
75%	28.05009	43.01642		
90%	31.49893	43.97919	Variance	39.75775
95%	33.98332	44.53403	Skewness	.1881963
99%	40.34642	45.80979	Kurtosis	3.048037

OLS:

Source	SS	df	MS	Number of obs	=	2,000
Model	52555.2814	3	17518.4271	F(3, 1996)	=	140.75
Residual	248439.676	1,996	124.468775	Prob > F	=	0.0000
Total	300994.957	1,999	150.572765	R-squared	=	0.1746
				Adj R-squared	=	0.1734
				Root MSE	=	11.157

wage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
age	.369376	.0324995	11.37	0.000	.3056395	.4331124
education	1.024154	.0863307	11.86	0.000	.8548468	1.193462
married	1.269777	.5790207	2.19	0.028	.1342283	2.405325
_cons	-11.7165	1.411936	-8.30	0.000	-14.48552	-8.947476

Tabla comparativa:

	(1) OLS	(2) Heckman (M~)	(3) Heckman (T~)
main			
age	0.369*** (0.0325)	0.212*** (0.0214)	0.211*** (0.0225)
education	1.024*** (0.0863)	0.988*** (0.0542)	0.980*** (0.0547)
married	1.270** (0.579)	0.0663 (0.376)	0.0864 (0.378)
_cons	-11.72*** (1.412)	0.497 (1.079)	0.730 (1.249)
dwage			
age		0.0364*** (0.00417)	0.0347*** (0.00423)
education		0.0556*** (0.0108)	0.0584*** (0.0110)
married		0.450*** (0.0727)	0.431*** (0.0742)
children		0.439*** (0.0278)	0.447*** (0.0287)
_cons		-2.489*** (0.190)	-2.467*** (0.193)
/			
athrho		0.875*** (0.102)	
lnsigma		1.793*** (0.0276)	
/mills			
lambda			4.021*** (0.613)
N	2000	2000	2000
R-sq	0.175		

Standard errors in parentheses
 * p<0.10, ** p<0.05, *** p<0.01

Ejercicio 3: Ecuación Salarial para las Mujeres II.

Conceptualmente, se va a repetir el ejercicio anterior utilizando la base de datos “mroz.dta” que ya se ha utilizado. Ahora, se pide modelar, explícitamente, la ecuación de selección con un Probit y la ecuación estructural con un modelo lineal aumentada por la inversa del ratio de Mills. Reportar el efecto marginal sobre las horas trabajadas, correctamente, estimado.

OLS:

Source	SS	df	MS	Number of obs	=	753
Model	119885614	6	19980935.6	F(6, 746)	=	33.05
Residual	451024110	746	604589.96	Prob > F	=	0.0000
Total	570909724	752	759188.463	R-squared	=	0.2100
				Adj R-squared	=	0.2036
				Root MSE	=	777.55

hours	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
kidsge6	-13.56954	23.87531	-0.57	0.570	-60.44032	33.30125
age	-17.10219	4.127445	-4.14	0.000	-25.20499	-8.999404
educ	23.9582	13.41096	1.79	0.074	-2.369512	50.28591
exper	74.12513	10.26049	7.22	0.000	53.98227	94.268
nwifeinc	-4.336964	2.633972	-1.65	0.100	-9.507843	.833916
expersq	-.9264192	.3349462	-2.77	0.006	-1.583968	-.2688699
_cons	656.2857	264.8041	2.48	0.013	136.4358	1176.136

Heckman (Two Step):

Heckman selection model -- two-step estimates
(regression model with sample selection)

Number of obs = 753
Selected = 428
Nonselected = 325

Wald chi2(6) = 26.17
Prob > chi2 = 0.0002

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
hours						
kidsge6	-83.74795	33.16153	-2.53	0.012	-148.7433	-18.75256
age	-2.839866	6.990271	-0.41	0.685	-16.54054	10.86081
educ	-63.81931	21.02964	-3.03	0.002	-105.0366	-22.60196
exper	6.070658	21.16833	0.29	0.774	-35.4185	47.55982
nwifeinc	4.458736	4.03176	1.11	0.269	-3.443369	12.36084
expersq	.1358569	.5265464	0.26	0.796	-.896155	1.167869
_cons	2477.33	425.3662	5.82	0.000	1643.627	3311.032
dhours						
kidsge6	.036005	.0434768	0.83	0.408	-.049208	.1212179
age	-.0528527	.0084772	-6.23	0.000	-.0694678	-.0362376
educ	.1309047	.0252542	5.18	0.000	.0814074	.180402
exper	.1233476	.0187164	6.59	0.000	.0866641	.1600311
nwifeinc	-.0120237	.0048398	-2.48	0.013	-.0215096	-.0025378
expersq	-.0018871	.0006	-3.15	0.002	-.003063	-.0007111
kidslt6	-.8683285	.1185223	-7.33	0.000	-1.100628	-.636029
_cons	.2700768	.508593	0.53	0.595	-.7267473	1.266901
/mills						
lambda	-621.8712	199.0294	-3.12	0.002	-1011.962	-231.7808
rho	-0.74244					
sigma	837.60041					

Tabla comparativa:

	(1) OLS	(2) Heckman (T~)
main		
kidsge6	-13.57 (23.88)	-83.75** (33.16)
age	-17.10*** (4.127)	-2.840 (6.990)
educ	23.96* (13.41)	-63.82*** (21.03)
exper	74.13*** (10.26)	6.071 (21.17)
nwifeinc	-4.337 (2.634)	4.459 (4.032)
expersq	-0.926*** (0.335)	0.136 (0.527)
_cons	656.3** (264.8)	2477.3*** (425.4)
dhours		
kidsge6		0.0360 (0.0435)
age		-0.0529*** (0.00848)
educ		0.131*** (0.0253)
exper		0.123*** (0.0187)
nwifeinc		-0.0120** (0.00484)
expersq		-0.00189*** (0.000600)
kidslt6		-0.868*** (0.119)
_cons		0.270 (0.509)
/mills		
lambda		-621.9*** (199.0)
N	753	753
R-sq	0.210	

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Efectos marginales (promedio) con censura en Heckman (Two Step):

Average marginal effects
Model VCE: Conventional

Number of obs = 753

Expression: $E(\text{hours}^* | \text{hours} > 0)$, $\text{predict}(\text{ystar}(0, .))$
dy/dx wrt: kidsge6 age educ exper nwifeinc expersq kidslt6

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
kidsge6		-81.38955	32.34639	-2.52	0.012	-144.7873	-17.99179
age		-2.759893	6.777967	-0.41	0.684	-16.04446	10.52468
educ		-62.02211	20.76646	-2.99	0.003	-102.7236	-21.32059
exper		5.899704	20.52631	0.29	0.774	-34.33112	46.13052
nwifeinc		4.333175	3.931451	1.10	0.270	-3.372327	12.03868
expersq		.132031	.5124218	0.26	0.797	-.8722971	1.136359
kidslt6		0	(omitted)				

Efectos marginales (promedio) con truncamiento en Heckman (Two Step):

Average marginal effects
Model VCE: Conventional

Number of obs = 753

Expression: $E(\text{hours} | \text{hours} > 0)$, $\text{predict}(e(0, .))$
dy/dx wrt: kidsge6 age educ exper nwifeinc expersq kidslt6

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
kidsge6		-73.14433	29.50712	-2.48	0.013	-130.9772	-15.31144
age		-2.4803	6.058449	-0.41	0.682	-14.35464	9.394042
educ		-55.73892	19.50195	-2.86	0.004	-93.96204	-17.5158
exper		5.302031	18.34814	0.29	0.773	-30.65967	41.26373
nwifeinc		3.894199	3.565572	1.09	0.275	-3.094194	10.88259
expersq		.1186556	.4620408	0.26	0.797	-.7869278	1.024239
kidslt6		0	(omitted)				

Efectos marginales (condicionales) con censura en Heckman (Two Step):

Conditional marginal effects
Model VCE: Conventional

Number of obs = 753

Expression: $E(\text{hours}^* | \text{hours} > 0), \text{predict}(\text{ystar}(0, .))$
 dy/dx wrt: kidsge6 age educ exper nwifeinc expersq kidslt6
 At: kidsge6 = 1.353254 (mean)
 age = 42.53785 (mean)
 educ = 12.28685 (mean)
 exper = 10.63081 (mean)
 nwifeinc = 20.12896 (mean)
 expersq = 178.0385 (mean)
 kidslt6 = .2377158 (mean)

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
kidsge6		-81.62997	32.48895	-2.51	0.012	-145.3071	-17.9528
age		-2.768046	6.79893	-0.41	0.684	-16.0937	10.55761
educ		-62.20532	20.85318	-2.98	0.003	-103.0768	-21.33383
exper		5.917131	20.58963	0.29	0.774	-34.4378	46.27207
nwifeinc		4.345974	3.9435	1.10	0.270	-3.383144	12.07509
expersq		.1324211	.5138982	0.26	0.797	-.8748009	1.139643
kidslt6		0	(omitted)				

Efectos marginales (condicionales) con truncamiento en Heckman (Two Step):

Conditional marginal effects
Model VCE: Conventional

Number of obs = 753

Expression: $E(\text{hours} | \text{hours} > 0), \text{predict}(e(0, .))$
 dy/dx wrt: kidsge6 age educ exper nwifeinc expersq kidslt6
 At: kidsge6 = 1.353254 (mean)
 age = 42.53785 (mean)
 educ = 12.28685 (mean)
 exper = 10.63081 (mean)
 nwifeinc = 20.12896 (mean)
 expersq = 178.0385 (mean)
 kidslt6 = .2377158 (mean)

		Delta-method				[95% conf. interval]	
		dy/dx	std. err.	z	P> z		
kidsge6		-73.52816	29.74524	-2.47	0.013	-131.8278	-15.22856
age		-2.493316	6.090065	-0.41	0.682	-14.42962	9.442992
educ		-56.03141	19.67987	-2.85	0.004	-94.60325	-17.45957
exper		5.329854	18.4439	0.29	0.773	-30.81952	41.47923
nwifeinc		3.914635	3.586477	1.09	0.275	-3.114731	10.944
expersq		.1192782	.4644865	0.26	0.797	-.7910986	1.029655
kidslt6		0	(omitted)				