# <u>Trabajo Práctico Nº 5:</u> Modelos para Variables Dependientes Limitadas - Tobit.

## **Ejercicio 1:** Variables Censuradas (Modelo Tobit I).

El modelo Tobit es relevante cuando la variable dependiente y de una regresión lineal se observa solo en algún intervalo de su soporte, porque, en este caso, los estimadores de MCC no son consistentes.

(a) Considerar la base "auto.dta". Estimar el modelo:

$$mpg = \alpha + \beta wgt + u$$
,

donde  $wgt = \frac{weight}{1000}$ . Luego, estimar el modelo generando una variable censurada suponiendo que no se observan autos con  $mpg \le 17$ . Estimar por MCC y utilizando un modelo Tobit. Comparar.

#### OLS:

Source	SS	df	MS	Number of ob F(1, 72)	-	74 134 62
	1591.99024	72		Prob > F	=	0.0000 0.6515 0.6467
Total	2443.45946					3.4389
mpg	Coefficient		t P	?> t  [95%	conf.	interval]
		.5178782	-11.60	0.000 -7.041 0.000 36.22		
OLS (ll(17)):						
Source	SS +	df	MS	Number of ob $F(1, 72)$		74 95.06
	1138.32073   862.219806		11.9752751	Prob > F	=	0.0000 0.5690 0.5630
Total	2000.54054					3.4605

mpg\_a | Coefficient Std. err. t P>|t| [95% conf. interval]

wgt | -5.080912 .5211373 -9.75 0.000 -6.11978 -4.042044 \_cons | 37.12539 1.62416 22.86 0.000 33.88769 40.3631

# <u>Tobit (ll(17)):</u>

Tobit regressi	Tobit regression				ber of obs	=	74
					Uncensore	d =	56
Limits: Lower	= 17			Left-censored =			
Upper	= +inf				Right-censore	d =	0
				LR	chi2(1)	=	72.85
				Pro	b > chi2	=	0.0000
Log likelihood			Pseudo R2			0.1815	
	Coefficient	Std. err.	t	P> t	[95% conf.	int	erval]
wat		.700257	-9.82	0.000	-8.268661	<b>-</b> -5	.47744
cons	41.49856	2.058384	20.16	0.000	37.3962	45	.60091
+							
<pre>var(e.mpg_a) </pre>	14.78942	2.817609			10.11698	21	.61977

## Tabla comparativa:

	(1) OLS	(2) OLS 11(17)	(3) Tobit 11(17)
main			
wgt	-6.009*** (0.518)	-5.081*** (0.521)	-6.873*** (0.700)
_cons	39.44*** (1.614)	37.13*** (1.624)	41.50*** (2.058)
/			
<pre>var(e.mpg_a)</pre>			14.79*** (2.818)
N P-sq	74 0.652	74 0.569	74
R-sq pseudo R-sq	0.052	0.309	0.182
Standard errors	in parentheses		

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

**(b)** Repetir el inciso anterior suponiendo que, ahora, no se observan autos con  $mpg \ge 24$ .

# OLS (ul(24)):

Source	SS	df	MS	Numb - F(1,	er of obs	=	74 186.15
Model Residual		1 72		1 Prob 4 R-sq	,	=	0.0000 0.7211 0.7172
Total	958		13.123287	_	-	=	1.9264
mpg_b	Coefficient	Std. err.		P> t	[95% cc	onf.	interval]
wgt   _cons	-3.958119 31.95138	.2901034 .9041273	-13.64 35.34	0.000	-4.53642 30.1490		-3.379808 33.75372

## <u>Tobit (ul(24)):</u>

Tobit regressi	on			Nur	mber of obs		74 51
Limits: Lower Upper					Uncensore Left-censore Right-censore	ed =	0 23
Log likelihood	= -129.8279			Pro	chi2(1) bb > chi2 eudo R2	=	90.72 0.0000 0.2589
mpg_b	Coefficient	Std. err.	t	P> t	[95% conf.	int	terval]
wgt     cons	-5.080645 36.08037				-5.947461 33.22628		.213829
var(e.mpg_b)	5.689927	1.166256			3.781783	8	.560846

## Tabla comparativa:

	(1) OLS	(2) OLS ul(24)	(3) Tobit ul(24)
main			
wgt	-6.009*** (0.518)	-3.958*** (0.290)	-5.081*** (0.435)
_cons	39.44*** (1.614)	31.95*** (0.904)	36.08*** (1.432)
/ var(e.mpg_b)			5.690*** (1.166)
			(1.100)
N R-sq	74 0.652	74 0.721	74
pseudo R-sq		· /21	0.259

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

(c) ¿Cómo se interpretan los coeficientes del modelo? Computar los efectos marginales.

Los coeficientes estimados miden cómo cambia la variable latente no observada con respecto a los cambios en las variables independientes, *céteris páribus*.

#### Efectos marginales (condicionales) con censura en Tobit (ll(17)):

```
Conditional marginal effects

Model VCE: OIM

Expression: E(mpg_a*|mpg_a>17), predict(ystar(17,.))

dy/dx wrt: wgt

1._at: wgt = 1

2._at: wgt = 2

3._at: wgt = 3

4._at: wgt = 4

| Delta-method | dy/dx std. err. z P>|z| [95% conf. interval]

wgt | at |

_at |

1 | -6.873035 | 1.389235 | -4.95 | 0.000 | -9.595886 | -4.150183 |

2 | -6.855268 | .7044715 | -9.73 | 0.000 | -8.236007 | -5.47453 |

3 | -5.797116 | .5880797 | -9.86 | 0.000 | -6.949731 | -4.644501 |

4 | -1.499391 | .3662326 | -4.09 | 0.000 | -2.217194 | -.7815884
```

#### Efectos marginales (condicionales) con truncamiento en Tobit (ll(17)):

### Efectos marginales (condicionales) con censura en Tobit (ul(24)):

```
Conditional marginal effects
                                                                                                           Number of obs = 74
Model VCE: OIM
Expression: E(mpg_b*|mpg_b<24), predict(ystar(.,24))</pre>
dy/dx wrt: wgt
1._at: wgt = 1
2._at: wgt = 2
3._{at: wgt = 3}
4. at: wgt = 4
                                   Delta-method
                                 dy/dx std. err.
                                                                          z P>|z| [95% conf. interval]
                     wgt
                _at |

    1
    | -.0085382
    .0114991
    -0.74
    0.458
    -.031076
    .0139997

    2
    | -1.069716
    .2842071
    -3.76
    0.000
    -1.626752
    -.5126807

    3
    | -4.610593
    .3715716
    -12.41
    0.000
    -5.33886
    -3.882326

    4
    | -5.079249
    .4349007
    -11.68
    0.000
    -5.931638
    -4.226859
```

#### Efectos marginales (condicionales) con truncamiento en Tobit (ul(24)):

```
Conditional marginal effects
                                              Number of obs = 74
Model VCE: OIM
Expression: E(mpg b|mpg b<24), predict(e(.,24))</pre>
dy/dx wrt: wgt
1._at: wgt = 1
2. at: wgt = 2
3._{at: wgt = 3}
4. at: wgt = 4
               Delta-method
                               z P>|z|
              dy/dx std.err.
                                           [95% conf. interval]
wgt
       _at |
       3 | -3.681238 .3548315 -10.37 0.000 -4.376695 -2.985781
```

4 | -5.06274 .4362475 -11.61 0.000 -5.917769 -4.20771

## **Ejercicio 2:** Variables Censuradas (Modelo Tobit II).

El siguiente ejercicio está tomado de Cameron & Trivedi. La variable dependiente para el gasto ambulatorio (ambulatory expenditure, ambexp) y los regresores (age, female, educ, blhisp, totchr, ins) se obtienen de la encuesta Medical Expenditure Panel Survey de 2001.

(a) Abrir y describir la base "mus16datav2.dta". ¿Qué se puede decir sobre el cumplimiento de las condiciones que requiere un Tobit?

V	ariable	Obs	Mean	Std. dev.	Min	Max
	ambexp	3,328	1386.519	2530.406	0	49960
	age	3,328	4.056881	1.121212	2.1	6.4
	female	3,328	.5084135	.5000043	0	1
	educ	3,328	13.40565	2.574199	0	17
	blhisp	3,328	.3085938	.4619824	0	1
	totchr	3,328	.4831731	.7720426	0	5
	ins	3,328	.3650841	.4815261	0	1
			ambexp			
	Percentiles	Sı	mallest			
1%	22		1			
5%	67		2			
10%	107		2	Obs		2,802
25%	275		4	Sum of wgt.		2,802
50%	779			Mean	1	646.8
			Largest	Std. dev.	267	8.914
75%	1913		28269			
90%	3967		30920	Variance	71	76579
95%	6027		34964	Skewness		99312
99%	12467		49960	Kurtosis		81969

Lo que se puede decir sobre el cumplimiento de las condiciones que requiere Tobit es que, en principio, la asimetría y la curtosis no normal (alejadas de 0 y 3, respectivamente) de la variable dependiente *ambexp* podrían deberse a regresores que están sesgados.

## Tobit:

Tobit regression Number of obs Uncensored						328 802
Limits: Lower = (	0			Left-censored	= .	526
Upper = +ini	f		R	ight-censored	=	0
			T.R. C	hi2(6)	= 694	0.7
				> chi2	= 0.0	
Log likelihood = -26359.424 Pseudo R2 =						130
ambexp   Coei	fficient Std. err	. t	P> t	[95% conf.	inter	val]
+						
age   31	14.1479 42.63366	7.37	0.000	230.557	397.	7388
female   68				502.9337		
educ				34.44865	107.	
blhisp   -				-734.7448	-325.	
totchr   12	244.578 60.51376	20.57	0.000	1125.93	1363	
ins   -10	67.4714 96.46088	-1.74		-356.6002	21.6	5734
_cons   -18	882.591 317.4305	-5.93	0.000	-2504.971	-1260	.212
var(e.ambexp)	6635296 179247.7	-		6292994	699	6217

### **(b)** Computar los efectos marginales.

## Efectos marginales (promedio) con censura en Tobit:

Average marginal effects Number of obs = 3,328

Model VCE: OIM

Expression: E(ambexp\*|ambexp>0), predict(ystar(0,.))

dy/dx wrt: age female educ blhisp totchr ins

	dy/dx	Delta-method std. err.	z	P> z	[95% conf.	interval]
age   female   educ   blhisp   totchr   ins	201.4409 439.2368 45.4411 -340.0509 798.06 -107.3876	27.29283 59.32556 11.89795 66.77218 38.00729 61.86227	7.38 7.40 3.82 -5.09 21.00 -1.74	0.000 0.000 0.000 0.000 0.000 0.083	147.9479 322.9608 22.12154 -470.922 723.5671 -228.6354	254.9338 555.5127 68.76066 -209.1799 872.5529 13.86024

### Efectos marginales (promedio) con truncamiento en Tobit:

Average marginal effects Number of obs = 3,328

Model VCE: OIM

Expression: E(ambexp|ambexp>0), predict(e(0,.)) dy/dx wrt: age female educ blhisp totchr ins

	dy/dx	Delta-method std. err.	Z	P> z	[95% conf.	interval]
age	147.796	20.14716	7.34	0.000	108.3083	187.2838
female	322.2656	43.7895	7.36	0.000	236.4397	408.0914
educ	33.33988	8.742173	3.81	0.000	16.20554	50.47422
blhisp	-249.4935	49.12834	-5.08	0.000	-345.7832	-153.2037
totchr	585.5322	29.01047	20.18	0.000	528.6727	642.3917
ins	-78.78967	45.40264	-1.74	0.083	-167.7772	10.19787

### Efectos marginales (condicionales) con censura en Tobit:

Conditional marginal effects Number of obs = 3,328

Model VCE: OIM

Expression: E(ambexp\*|ambexp>0), predict(ystar(0,.))

dy/dx wrt: age female educ blhisp totchr ins At: age = 4.056881 (mean)

At: age = 4.056881 (mean) female = .5084135 (mean) educ = 13.40565 (mean) blhisp = .3085938 (mean) totchr = .4831731 (mean) ins = .3650841 (mean)

	dy/dx	Delta-method std. err.	z	P> z	[95% conf.	interval]
age	207.526	28.2054	7.36	0.000	152.2444	262.8076
female	452.5052	61.30341	7.38	0.000	332.3528	572.6577
educ	46.81378	12.26552	3.82	0.000	22.77381	70.85375
blhisp	-350.3232	68.86825	-5.09	0.000	-485.3025	-215.3439
totchr	822.1678	40.61039	20.25	0.000	742.5729	901.7627
ins	-110.6315	63.74577	-1.74	0.083	-235.5709	14.30787

#### Efectos marginales (condicionales) con truncamiento en Tobit:

(c) Computar los efectos marginales haciendo las cuentas con los comandos de escalares y matrices de Stata.

#### Stata.

(d) Considerar la variable dependiente en logaritmos. ¿Qué interpretación tiene esto sobre la variable dependiente? ¿Qué complicaciones introduce en el análisis? Estimar un Tobit para el logaritmo de ambexp.

La variable dependiente en logaritmos introduce dos complicaciones en el análisis: un umbral distinto de cero y una variable dependiente lognormal.

# OLS (con variable dependiente en logaritmos):

Source	SS S	df	MS	Number F(6, 3	of obs	=	3,328 169.68
Model Residual	5772.79592		962.132653 5.67028725	Prob > R-squa	F	= =	0.0000 0.2346 0.2332
Total	24603.8199	3 <b>,</b> 327	7.39519683	Root M	-	=	2.3812
lambexp	Coefficient	Std. err.	t 1	 P> t  	[95% con:	f.	interval]
age female educ blhisp totchr ins _cons	1.144695	.038348 .0833418 .0165414 .0928854 .0553699 .0869061 .2812597	13.73 ( 6.90 ( -7.90 ( 19.13 ( 2.39 (	0.000 0.000 0.000 0.000 0.000 0.017 0.000	.2495436 .9812886 .0816757 9162938 .9508324 .0374394 1.177304		.3999199 1.308102 .1465403 5520571 1.167958 .3782293 2.280224

# <u>Tobit (con variable dependiente en logaritmos):</u>

Tobit regression	ı			Numbe	er of obs Uncensored	•
Limits: Lower = Upper =					Left-censored lght-censored	= 526
Log likelihood =	-7494.29				ni2(6) > chi2 do R2	= 831.03 = 0.0000 = 0.0525
lambexp	Coefficient	Std. err.	t	P> t	[95% conf	. interval]
	.138446 8731611 1.161268	.0453222 .0986074 .0196568 .1102504 .0649655 .102613 .3350343	8.01 13.61 7.04 -7.92 17.88 2.55 2.76	0.000 0.000 0.000 0.000 0.000 0.011 0.006	-1.089327 1.033891	1.535146 .1769866 6569955 1.288644
var(e.lambexp)	7.735265	.2181984			7.319064	8.175133

# <u>Tabla comparativa:</u>

	(1) OLS (log)	(2) Tobit (log)
main age	0.325***	0.363***
female	1.145*** (0.0833)	1.342*** (0.0986)
educ	0.114*** (0.0165)	0.138*** (0.0197)
blhisp	-0.734*** (0.0929)	-0.873*** (0.110)
totchr	1.059*** (0.0554)	1.161*** (0.0650)
ins	0.208** (0.0869)	0.261** (0.103)
_cons	1.729*** (0.281)	0.924*** (0.335)
/ var(e.lamb~)		7.735*** (0.218)
N	3328	3328
R-sq pseudo R-sq	0.235	0.053
Standard errors	in parentheses	

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

## **Ejercicio 3:** Variables Censuradas (Modelo Tobit III).

Considerar la base de datos "mroz.dta", que posee datos que permiten estudiar la oferta laboral anual de mujeres casadas. Considerar las horas trabajadas, hours, y las explicativas nwifeinc, educ, exper, expersq, age, kidslt6, kidsge6. Estimar un modelo lineal y un modelo Tobit. Comparar. Computar los efectos marginales.

### OLS:

Source	SS	df	MS		ber of obs	=	753
Model   Residual	151647606 419262118	7 745	21663943.7 562767.944	Pro R-s	, 745) b > F quared	=	38.50 0.0000 0.2656
Total	570909724	752	759188.463	_	R-squared t MSE	=	0.2587 750.18
hours	Coefficient	Std. err.	t	 P> t	[95% cor	nf.	interval]
kidslt6   kidsge6   age   educ   exper   nwifeinc   expersq   _cons	-442.0899 -32.77923 -30.51163 28.76112 65.67251 -3.446636 7004939 1330.482	58.8466 23.17622 4.363868 12.95459 9.962983 2.544 .3245501 270.7846	-1.41 -6.99 2.22 6.59 -1.35 -2.16	0.000 0.158 0.000 0.027 0.000 0.176 0.031 0.000	-557.6148 -78.2777 -39.07858 3.329283 46.11365 -8.440898 -1.337635 798.8906	7 3 3 5 5	-326.565 12.71924 -21.94469 54.19297 85.23138 1.547626 0633524 1862.074

### Tobit:

Tobit regressi	on			Nui	mber of obs Uncensore	
Limits: Lower Upper					Left-censore Right-censore	ed = 325
Log likelihood	= -3819.0946			Pro	chi2(7) ob > chi2 eudo R2	= 271.59 = 0.0000 = 0.0343
hours	Coefficient	Std. err.	t	P> t	[95% conf.	interval]
age   educ	-16.21805 -54.40491 80.64541 131.564 -8.814226	38.6413 7.418483 21.58318 17.27935 4.459089 .5376606	-7.99 -0.42 -7.33 3.74 7.61 -1.98 -3.47 2.16	0.675 0.000 0.000 0.000 0.048	-92.07668 -68.9685 38.27441	59.64057 -39.84133 123.0164 165.48606037068086455
var(e.hours)	1258927	93304.48			1088458	1456093

# <u>Tabla comparativa:</u>

(58.85) (111.9) kidsge6 -32.78		(1) OLS	(2) Tobit
(23.18) (38.64)  age			-894.0*** (111.9)
(4.364) (7.418)  educ 28.76** 80.65*** (12.95) (21.58)  exper 65.67*** 131.6*** (9.963) (17.28)  nwifeinc -3.447 -8.814** (2.544) (4.459)  expersq -0.700** -1.864*** (0.325) (0.538)  _cons 1330.5*** 965.3** (270.8) (446.4)  / var(e.hours) 1258926.8*** (93304.5)	kidsge6		
(12.95) (21.58)  exper 65.67*** 131.6*** (9.963) (17.28)  nwifeinc -3.447 -8.814** (2.544) (4.459)  expersq -0.700** -1.864*** (0.325) (0.538)  _cons 1330.5*** 965.3** (270.8) (446.4)  / var(e.hours) 1258926.8*** (93304.5)	age		
(9.963) (17.28)  nwifeinc -3.447 -8.814** (2.544) (4.459)  expersq -0.700** -1.864*** (0.325) (0.538)  _cons 1330.5*** 965.3** (270.8) (446.4)  / var(e.hours) 1258926.8*** (93304.5)	educ		80.65*** (21.58)
(2.544) (4.459)  expersq -0.700** -1.864*** (0.325) (0.538)  _cons 1330.5*** 965.3** (270.8) (446.4)  / var(e.hours) 1258926.8*** (93304.5)	exper		131.6*** (17.28)
(0.325) (0.538)  _cons	nwifeinc		
(270.8) (446.4)  / var(e.hours) 1258926.8***	expersq		
var(e.hours) 1258926.8*** (93304.5)	_cons		
N 753 753	,		1258926.8*** (93304.5)
R-sq 0.266	N R-sa	753 0 266	753
pseudo R-sq 0.200 0.034	_	0.200	0.034

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

#### Efectos marginales (promedio) con censura en Tobit:

Average marginal effects Number of obs = 753

Model VCE: OIM

Expression: E(hours\*|hours>0), predict(ystar(0,.))

dy/dx wrt: kidslt6 kidsge6 age educ exper nwifeinc expersq

	dy/dx	Delta-method std. err.	Z	P> z	[95% conf.	interval]
kidslt6   kidsge6   age   educ   exper   nwifeinc   expersq	-526.2776	64.70619	-8.13	0.000	-653.0994	-399.4558
	-9.546986	22.75224	-0.42	0.675	-54.14056	35.04659
	-32.02622	4.29211	-7.46	0.000	-40.4386	-23.61384
	47.47306	12.6214	3.76	0.000	22.73558	72.21054
	77.44703	9.99765	7.75	0.000	57.85199	97.04206
	-5.188619	2.621409	-1.98	0.048	-10.32649	0507514
	-1.09736	.3155945	-3.48	0.001	-1.715914	4788063

#### Efectos marginales (promedio) con truncamiento en Tobit:

Average marginal effects Number of obs = 753

Model VCE: OIM

Expression: E(hours|hours>0), predict(e(0,.))

dy/dx wrt: kidslt6 kidsge6 age educ exper nwifeinc expersq

| Delta-method | dy/dx std. err. z P>|z| [95% conf. interval] | | kidslt6 | -402.5505 | 50.74874 | -7.93 | 0.000 | -502.0162 | -303.0848 | kidsge6 | -7.302504 | 17.40426 | -0.42 | 0.675 | -41.41423 | 26.80922 | age | -24.4969 | 3.362491 | -7.29 | 0.000 | -31.08726 | -17.90654 | educ | 36.31221 | 9.703035 | 3.74 | 0.000 | 17.29461 | 55.32981 | exper | 59.23934 | 7.83368 | 7.56 | 0.000 | 43.88561 | 74.59308 | nwifeinc | -3.968782 | 2.007582 | -1.98 | 0.048 | -7.903569 | -.0339945 | expersq | -.8393724 | .2423183 | -3.46 | 0.001 | -1.314307 | -.3644373

#### Efectos marginales (condicionales) con censura en Tobit:

```
Conditional marginal effects
                                                                         Number of obs = 753
Model VCE: OIM
Expression: E(hours*|hours>0), predict(ystar(0,.))
dy/dx wrt: kidslt6 kidsge6 age educ exper nwifeinc expersq
At: kidslt6 = .2377158  (mean)
    kidsge6 = 1.353254  (mean)
    age = 42.53785 (mean)
educ = 12.28685 (mean)
    exper = 10.63081 (mean)
    nwifeinc = 20.12896 (mean)
     expersq = 178.0385 (mean)
                              Delta-method
                       dy/dx std.err.
                                                   z P>|z|
                                                                      [95% conf. interval]
               kidslt6 | -540.2567 66.62387 -8.11 0.000 -670.8371 -409.6763
kidsge6 | -9.800576 23.36132 -0.42 0.675 -55.58792 35.98677
age | -32.87691 4.457699 -7.38 0.000 -41.61384 -24.13998
educ | 48.73405 12.9634 3.76 0.000 23.32625 74.14185
exper | 79.50419 10.30495 7.72 0.000 59.30685 99.70153
```

#### Efectos marginales (condicionales) con truncamiento en Tobit:

Conditional marginal effects Number of obs = 753

nwifeinc | -5.32644 2.690724 -1.98 0.048 -10.60016 -.0527175 expersq | -1.126508 .3232603 -3.48 0.000 -1.760087 -.49293

Model VCE: OIM

Expression: E(hours|hours>0), predict(e(0,.))

dy/dx wrt: kidslt6 kidsge6 age educ exper nwifeinc expersq

At: kidslt6 = .2377158 (mean) kidsge6 = 1.353254 (mean) age = 42.53785 (mean) educ = 12.28685 (mean) exper = 10.63081 (mean) nwifeinc = 20.12896 (mean) expersg = 178.0385 (mean)

	dy/dx 	Delta-method std. err.	Z	P> z	[95% conf.	interval]
kidslt6   kidsge6   age   educ   exper   nwifeinc   expersg	-379.9678 -6.892841 -23.12265 34.27513 55.91608 -3.746137	46.79714 16.42951 3.130037 9.117076 7.239109 1.89236	-8.12 -0.42 -7.39 3.76 7.72 -1.98	0.000 0.675 0.000 0.000 0.000 0.048	-471.6885 -39.09409 -29.25741 16.40599 41.72769 -7.455095 -1.237871	-288.2471 25.3084 -16.98789 52.14427 70.10447 03718