22) Tobit Model

Vitor Kamada

August 2018

The Tobit Model for Corner Solution Responses

$$y^* = x\beta + u, \quad u|x \sim N(0, \sigma^2)$$

$$y = max(0, y^*)$$

$$P(y = 0|x) = P(y^* < 0|x) = P(u < -x\beta|x) =$$
$$P(\frac{u}{\sigma} < -\frac{x\beta}{\sigma}|x) = \Phi(\frac{-x\beta}{\sigma}) = 1 - \Phi(\frac{x\beta}{\sigma})$$

$$\ell_i(\beta, \sigma) = 1(y_i = 0)log[1 - \Phi(\frac{x_i\beta}{\sigma})]$$
$$+1(y_i > 0)log\{(\frac{1}{\sigma}\phi(\frac{y_i - x_i\beta}{\sigma})\}$$

Interpreting the Tobit Estimates

$$E(y|x) = P(y > 0|x) \cdot E(y|y > 0, x)$$

$$= \Phi(\frac{x_i\beta}{\sigma})[x\beta + \sigma\lambda(\frac{x_i\beta}{\sigma})]$$

$$= \Phi(\frac{x_i\beta}{\sigma})x\beta + \sigma\phi(\frac{x_i\beta}{\sigma})$$

$$\lambda(c) = \frac{\phi(c)}{\Phi(c)}$$

Partial Effects

$$E(y|x) = \Phi(\frac{x_i\beta}{\sigma})x\beta + \sigma\phi(\frac{x_i\beta}{\sigma})$$

$$\frac{\partial E(y|x)}{\partial x_j} = \beta_j\Phi(\frac{x_i\beta}{\sigma})$$

PEA: $\beta_j\Phi(\frac{\bar{x}\hat{\beta}}{\hat{\sigma}})$

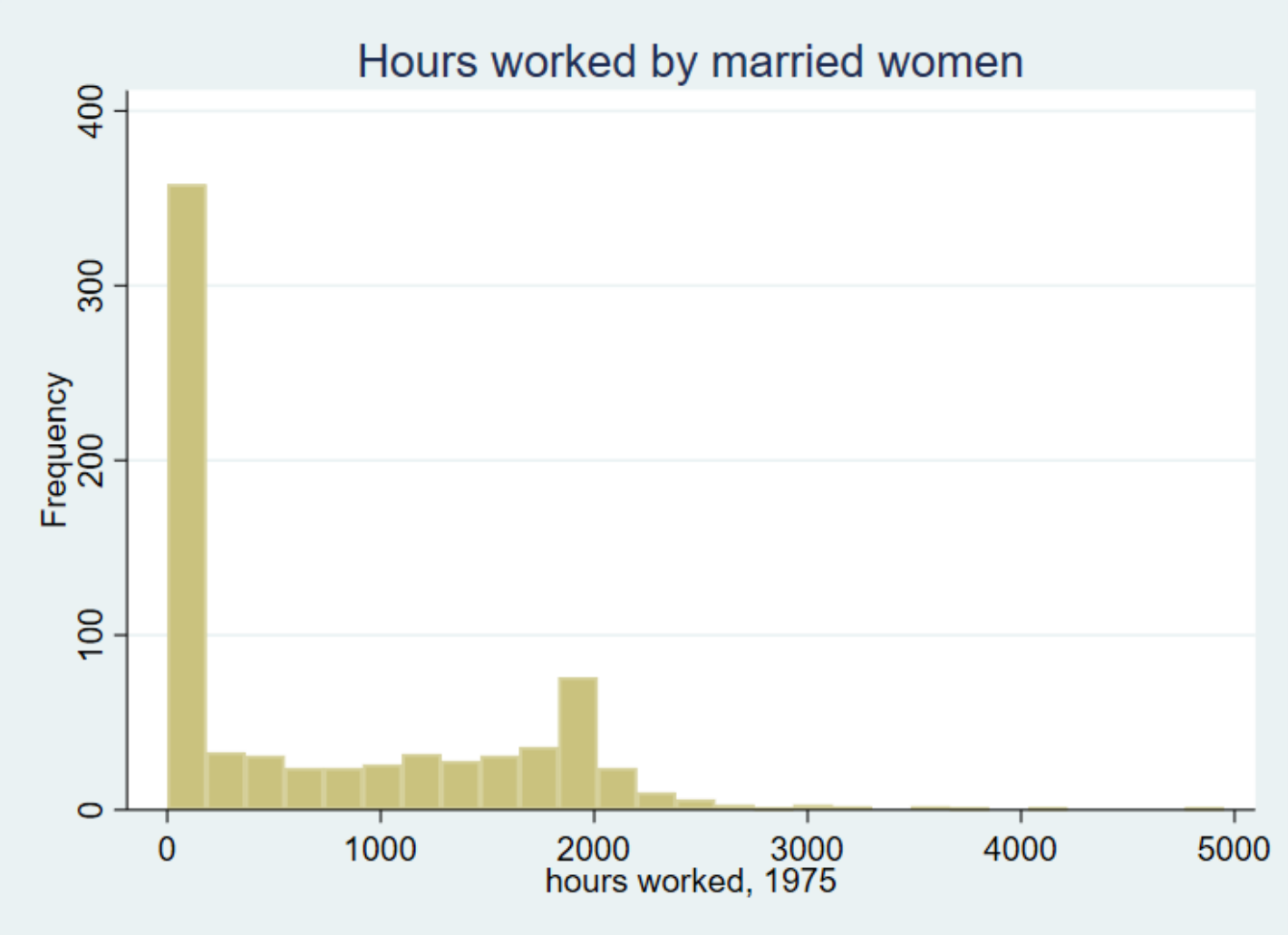
APE: $\beta_j[n^{-1}\sum_{i=1}^{n}\Phi(\frac{x_i\hat{\beta}}{\hat{\sigma}})]$

Annual Hours Equation for Married Women

Mroz (1987)

Variable	Obs	Mean	Std. Dev.	Min
hours	753	740.5764	871.3142	0
nwifeinc	753	20.12896	11.6348	-.0290575
educ	753	12.28685	2.280246	5
exper	753	10.63081	8.06913	0
expersq	753	178.0385	249.6308	0
age	753	42.53785	8.072574	30
kidslt6	753	.2377158	.523959	0
kidsge6	753	1.353254	1.319874	0

histogram hours, frequency title(Hours worked by married women)



reg hours nwifeinc educ exper expersq age kidslt6 kidsge6, vce(robust)

Linear regression

Number of obs	=	753	
F(7, 745)	=	45.81	
Prob > F	=	0.0000	
R-squared	=	0.2656	
Root MSE	=	750.18	

hours	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
nwifeinc	-3.446636	2.240662	-1.54	0.124	-7.845398	.9521268
educ	28.76112	13.03905	2.21	0.028	3.163468	54.35878
exper	65.67251	10.79419	6.08	0.000	44.48186	86.86316
expersq	-.7004939	.3720129	-1.88	0.060	-1.430812	.0298245
age	-30.51163	4.244791	-7.19	0.000	-38.84481	-22.17846
kidslt6	-442.0899	57.46384	-7.69	0.000	-554.9002	-329.2796
kidsge6	-32.77923	22.80238	-1.44	0.151	-77.5438	11.98535
_cons	1330.482	274.8776	4.84	0.000	790.8556	1870.109

tobit hours nwifeinc educ exper expersq age kidslt6 kidsge6, ll(0)

Tobit regression

Limits: lower = 0
upper = +inf

Log likelihood = -3819.0946

Number of obs	= 753
Uncensored	= 428
Left-censored	= 325
Right-censored	= 0
LR chi2(7)	= 271.59
Prob > chi2	= 0.0000
Pseudo R2	= 0.0343

hours	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
nwifeinc	-8.814226	4.459089	-1.98	0.048	-17.56808	-.0603706
educ	80.64541	21.58318	3.74	0.000	38.27441	123.0164
exper	131.564	17.27935	7.61	0.000	97.64211	165.486
expersq	-1.864153	.5376606	-3.47	0.001	-2.919661	-.8086455
age	-54.40491	7.418483	-7.33	0.000	-68.9685	-39.84133
kidslt6	-894.0202	111.8777	-7.99	0.000	-1113.653	-674.3875
kidsge6	-16.21805	38.6413	-0.42	0.675	-92.07668	59.64057
_cons	965.3068	446.4351	2.16	0.031	88.88827	1841.725
var(e.hours)	1258927	93304.48			1088458	1456093

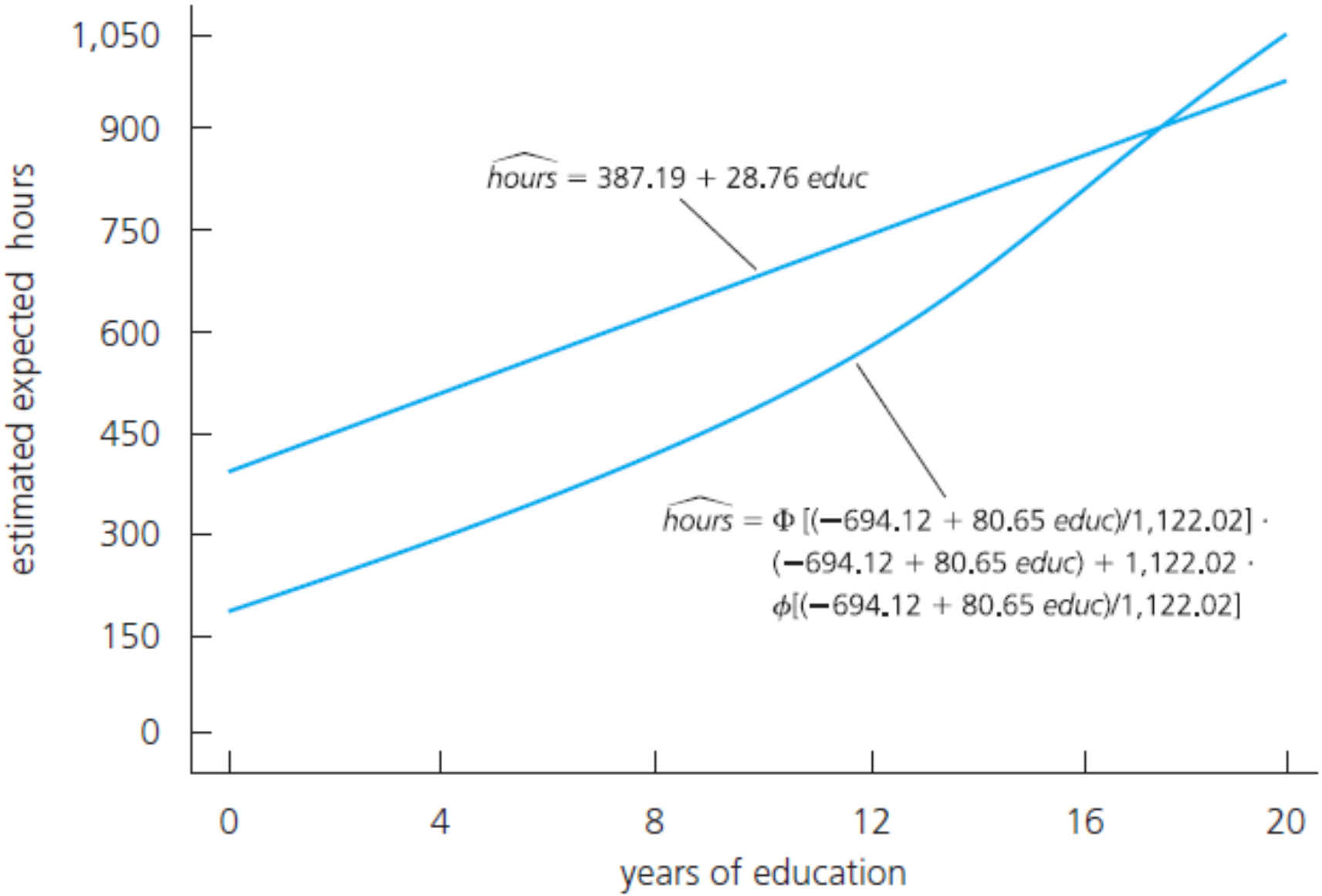
margins, dydx(*) predict(ystar(0,.))

Average marginal effects — Number of obs = 753
Model VCE : **OIM**

Expression : **E(hours*|hours>0), predict(ystar(0,.))**
dy/dx w.r.t. : **nwifeinc educ exper expersq age kidslt6 kidsge6**

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
nwifeinc	-5.188619	2.621409	-1.98	0.048	-10.32649	-.0507515
educ	47.47306	12.6214	3.76	0.000	22.73558	72.21054
exper	77.44703	9.99765	7.75	0.000	57.85199	97.04206
expersq	-1.09736	.3155945	-3.48	0.001	-1.715914	-.4788063
age	-32.02622	4.292111	-7.46	0.000	-40.4386	-23.61384
kidslt6	-526.2776	64.70619	-8.13	0.000	-653.0994	-399.4558
kidsge6	-9.546986	22.75224	-0.42	0.675	-54.14056	35.04659

Wooldridge (2016)



reg nwifeinc huseduc educ exper expersq age kidslt6 kidsge6

nwifeinc	Coef.	Std. Err.	t	P>|t|
huseduc	1.178155	.1609449	7.32	0.000
educ	.6746951	.2136829	3.16	0.002
exper	-.3129877	.1382549	-2.26	0.024
expersq	-.0004776	.0045196	-0.11	0.916
age	.3401521	.0597084	5.70	0.000
kidslt6	.8262719	.8183785	1.01	0.313
kidsge6	.4355289	.3219888	1.35	0.177
_cons	-14.72048	3.787326	-3.89	0.000

predict vhat, resid

tobit hours nwifeinc vhat educ exper expersq age kidslt6 kidsge6, ll(0)

hours	Coef.	Std. Err.	t	P>|t|
nwifeinc	-31.48209	16.03758	-1.96	0.050
vhat	24.41828	16.5845	1.47	0.141
educ	116.7811	32.75973	3.56	0.000
exper	124.3485	17.87499	6.96	0.000
expersq	-1.897196	.5371606	-3.53	0.000
age	-46.89236	8.957659	-5.23	0.000
kidslt6	-867.9116	112.9022	-7.69	0.000
kidsge6	-6.326126	39.16555	-0.16	0.872
_cons	722.1052	475.6883	1.52	0.129
var(e.hours)	1254045	92931.19		

Appendix I

$$z \sim N(0,1)$$

$$E(z|z > c) = \int_c^{\infty} \{z\phi(c)/[1-\Phi(c)]\}dz$$

$$= \int_c^{\infty} \frac{z}{\sqrt{2\pi}} exp(\frac{-z^2}{2})dz/[1-\Phi(c)]$$

$$= \int_c^{\infty} \frac{\partial}{\partial z}(-\frac{1}{\sqrt{2\pi}} exp(\frac{-z^2}{2}))dz/[1-\Phi(c)]$$

$$= [-\frac{1}{\sqrt{2\pi}} exp(\frac{-z^2}{2})]_c^{\infty}/[1-\Phi(c)]$$

$$= \frac{\phi(c)}{1-\Phi(c)}$$

Appendix II

If $z \sim N(0,1)$, then $E(z|z>0) = \frac{\phi(c)}{1-\Phi(c)}$

$$\phi(-c) = \phi(c) \text{ and } 1 - \Phi(-c) = \Phi(c)$$

$$E(y|y>0,x)$$

$$= x\beta + E(u|u > -x\beta)$$

$$= x\beta + \sigma E[(\tfrac{u}{\sigma})|(\tfrac{u}{\sigma}) > \tfrac{-x\beta}{\sigma}]$$

$$= x\beta + \sigma \frac{\phi(\frac{x_i\beta}{\sigma})}{\Phi(\frac{x_i\beta}{\sigma})}$$

Appendix III

$$\frac{d\Phi}{dc} = \phi(c) \text{ and } \frac{d\phi}{dc} = -c\phi(c)$$

$$E(y|x) = \Phi(\frac{x\beta}{\sigma})x\beta + \sigma\phi(\frac{x\beta}{\sigma})$$

$$\frac{\partial E(y|x)}{\partial x_j} = \frac{\beta_j}{\sigma}\Phi' x\beta + \beta_j\Phi(\cdot) + \frac{\beta_j}{\sigma}\sigma\phi'$$

$$= \frac{\beta_j}{\sigma}\phi x\beta + \beta_j\Phi(\cdot) - \beta_j\frac{x\beta}{\sigma}\phi$$

$$= \beta_j\Phi(\frac{x\beta}{\sigma})$$