Censored Regression Model Formulation

The usual index model formulation is

$$\cancel{y_i \le y_i^* \text{ if } y_i}$$

$$y_i^* = x_i'\beta + \epsilon_i$$

$$\epsilon_i \sim \text{iid } N(0, \sigma^2)$$

and

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \le 0 \end{cases}$$

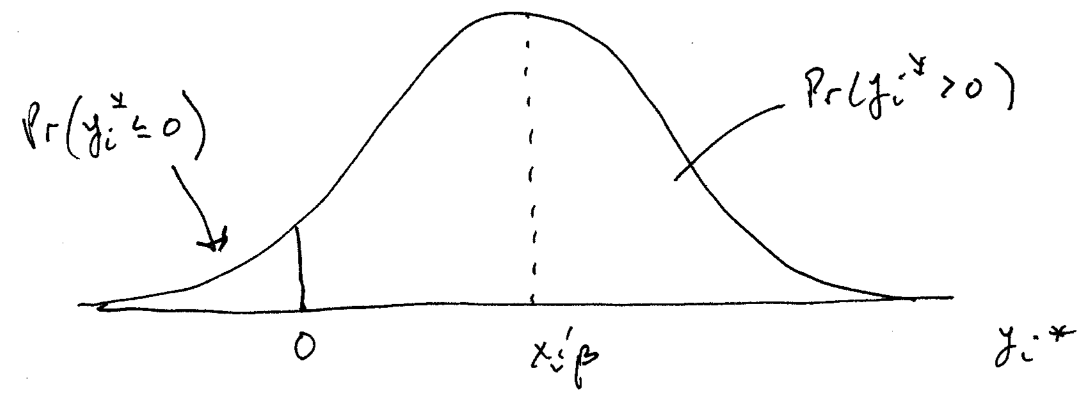
Alternatively, we could also have

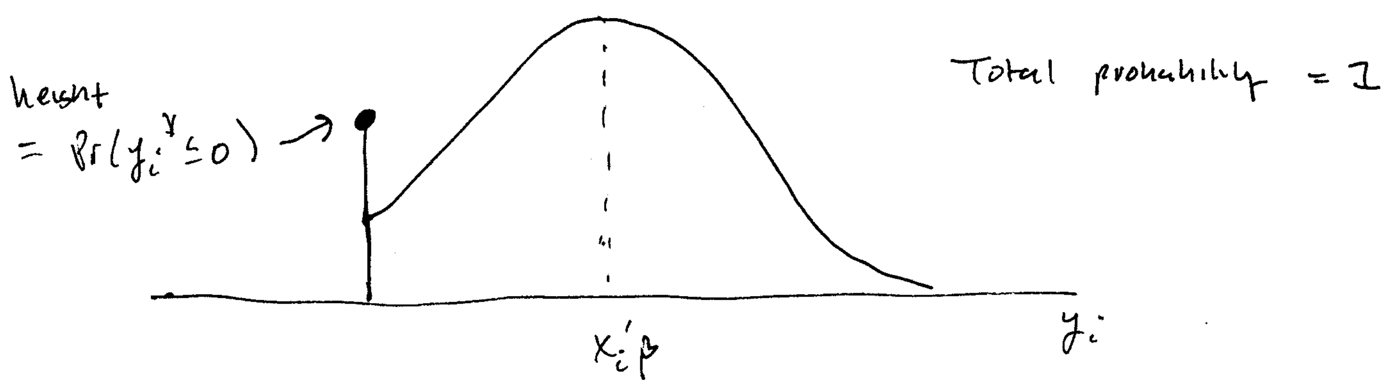
$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > \tau \\ \tau & \text{if } y_i^* \le \tau \end{cases}$$

or even

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > \tau_1 \\ \tau_2 & \text{if } y_i^* \le \tau_2 \end{cases}$$

Censored Normal Density Function





y_i has pdf $f(y_i^*; x_i, \beta, \sigma^2) = N(x_i'\beta, \sigma^2)$
for $y_i^* > 0$

y_i has pmf $Pr(y_i^* \le 0)$ for $y_i^* \le 0$

<u>Note:</u> $Pr(y_i^* \le 0) = Pr(x_i'\beta + \epsilon_i \le 0)$

$= Pr(\epsilon_i \le -x_i'\beta)$

$= Pr\left(\frac{\epsilon_i}{\sigma} \le -\frac{x_i'\beta}{\sigma}\right)$

$= \Phi\left(-\frac{x_i'\beta}{\sigma}\right)$

$= 1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)$

Hence

$$f(y_i | x_i, \beta, \sigma^2) = \frac{1}{\sigma}\varphi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \qquad y_i^* > 0$$

$$= 1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right) \qquad y_i = 0$$

This is a mixture of continuous & discrete densities.

Given an iid sample, the likelihood function is

$$L(\beta, \sigma^2 | y, x) = \prod_{\{y_i > 0\}} \frac{1}{\sigma}\varphi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \cdot \prod_{\{y_i = 0\}} \left(1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right)$$

and the log-likelihood is

$$\ln L(\beta, \sigma^2 | y, x) = \sum_{\{y_i > 0\}} -\frac{1}{2}\left\{\ln(2\pi) + \ln(\sigma^2) + \frac{(y_i - x_i'\beta)^2}{\sigma^2}\right]$$

$$+ \sum_{\{y_i = 0\}} \ln\left(1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right)$$

Amemiya (1973) Ecta proves that this likelihood satisfies all the usual regularity conditions.

Interpretation of the Censored Regression model

3 Conditional means of interest

$E[y_i^* | x_i]$: latent variable

$E[y_i | x_i]$: observed data

$E[y_i | x_i, y_i > 0]$: Non censored data

and each of these produce different sets of marginal effects.

(1) $$E[y_i^* | x_i] = x_i'\beta$$

$$\frac{\partial E[y_i^* | x_i]}{\partial x_i} = \beta$$

(2) $$E[y_i | x_i, y_i > 0] = x_i'\beta + \sigma\lambda\left(\frac{-x_i'\beta}{\sigma}\right)$$

$$\frac{\partial E[y_i | x_i, y_i > 0]}{\partial x_i} = \beta\left(1 - \delta\left(\frac{-x_i'\beta}{\sigma}\right)\right)$$

(3) Q: What is $E[y_i | x_i]$?

$$Pr\left(\frac{\epsilon_i}{\sigma} > -\frac{x_i'\beta}{\sigma}\right) = 1 - \Phi\left(-\frac{x_i'\beta}{\sigma}\right)$$

$$\lambda = \frac{\varphi}{1-\Phi} = \frac{\varphi(-x_i'\beta)}{1-\Phi(-\frac{x_i'\beta}{\sigma})} = \frac{\varphi(x_i)}{\Phi(x_i)}$$

Use result that

$$E[y_i|x_i] = E[y_i|x_i, y_i=0] \cdot Pr(y_i=0)$$

$$+ E[y_i|x_i, y_i>0] \cdot Pr(y_i>0)$$

$$= 0 + \left(x_i'\beta + \sigma\lambda\left(\frac{-x_i'\beta}{\sigma}\right)\right) * Pr(x_i'\beta + \epsilon_i$$

$$= \left(x_i'\beta + \sigma\lambda\left(\frac{-x_i'\beta}{\sigma}\right)\right)\left(1-\Phi\left(-\frac{x_i'\beta}{\sigma}\right)\right)$$

$$= \cancel{x_i'\beta}$$

Now use result that

$$\lambda\left(\frac{-x_i'\beta}{\sigma}\right) = \frac{\varphi(-\frac{x_i'\beta}{\sigma})}{1-\Phi(-\frac{x_i'\beta}{\sigma})} = \frac{\varphi(\frac{x_i'\beta}{\sigma})}{\Phi(\frac{x_i'\beta}{\sigma})}$$

$$1 - \Phi\left(-\frac{x_i'\beta}{\sigma}\right) = \Phi\left(\frac{x_i'\beta}{\sigma}\right)$$

to give

$$E[y_i|x_i] = \left\{x_i'\beta + \sigma\frac{\varphi(\frac{x_i'\beta}{\sigma})}{\Phi(\frac{x_i'\beta}{\sigma})}\right\} * \Phi\left(\frac{x_i'\beta}{\sigma}\right)$$

Expanding the expression gives

$$E[y_i|x_i] = x_i'\beta\,\Phi\left(\frac{x_i'\beta}{\sigma}\right) + \sigma\,\varphi\left(\frac{x_i'\beta}{\sigma}\right).$$

Now,

$$\frac{\partial E[y_i|x_i]}{\partial x_i} = \frac{\partial}{\partial x_i}(x_i'\beta)\cdot\Phi\left(\frac{x_i'\beta}{\sigma}\right) + x_i'\beta\,\frac{\partial}{\partial x_i}\Phi\left(\frac{x_i'\beta}{\sigma}\right)$$

$$+\ \sigma\,\frac{\partial}{\partial x_i}\varphi\left(\frac{x_i'\beta}{\sigma}\right)$$ via the chain rule.

$$\boxed{\begin{array}{l}\text{use fact}\\ \text{that}\\ \frac{d}{dx}\varphi(x) = -x\varphi(x)\\ \frac{d}{dx}\Phi(x) = \varphi(x)\end{array}} = \beta\,\Phi\left(\frac{x_i'\beta}{\sigma}\right) + x_i'\beta * \varphi\left(\frac{x_i'\beta}{\sigma}\right) * \frac{\beta}{\sigma}$$

$$+\ \cancel{\beta}\left(-\frac{x_i'\beta}{\cancel{\sigma}}\right)\varphi\left(\frac{x_i'\beta}{\sigma}\right) * \frac{\beta}{\sigma}$$

$$= \beta\,\Phi\left(\frac{x_i'\beta}{\sigma}\right) + 0$$

$$= \beta\,\Phi\left(\frac{x_i'\beta}{\sigma}\right)$$

$$< \beta$$

since $\Phi\left(\frac{x_i'\beta}{\sigma}\right)$ is between 0 and 1.

Example : # of hours worked by women

Quester & Greene "Divorce Risk and Wives' Labor supply Behavior", Social Science Quarterly, 1982

y_i = # of hours worked in survey year

Tobit Model Estimated by MLE

Variable	$\hat{\beta}_{MLE}$	$\partial E[y_i \mid x_i]/\partial x_i = \beta \cdot \Phi\left(\frac{x_i'\beta}{\sigma}\right)$
Constant	−1803.13	
t-ratio:	(−8.64)	
Small kids	−1324.84	−385.89
	(−19.78)	
Education Diff b/w Husbands & Wives	−48.08	−14.00
	(−4.77)	
Relative wage	312.07	90.90
	(5.71)	
2nd Marriage	175.85	51.51
	(3.47)	
Mean Divorce Prob	417.39	121.58
	(6.52)	
High Divorce Prob.	670.22	195.22
	(8.40)	

<u>Example</u> Fair "Theory of Extramarital Affairs"
JPE 1978

y_i = # of extramarital sexual encounters

y_i^* = utility index associated with affairs
(# of desired sexual encounters)

Tobit model

$$y_i = y_i^* \quad \text{if} \quad y_i^* > 0$$
$$= 0 \qquad\qquad y_i^* \leq 0$$

Variable	Psychology Today $\hat{\beta}_{MLE}$	Psychology Today t-stat	Redbook $\hat{\beta}_{MLE}$	Redbook t-stat
Constant	7.60	1.92	7.85	11
Occupation	0.213	0.67	0.314	3.77
Education	0.025	0.11	-0.0853	-2.34
Husband's Occupation			0.015	0.27
Marital Happiness	-2.27	-5.48	-1.53	-20
Age	-0.193	-2.37	-0.107	-4.24
# of years married	0.533	3.63	0.130	4.82
Children	1.02	0.79	-0.028	-0.36
Degree of Religiosity	-1.70	-4.15	-0.94	-11.03
Sex	0.95	0.88		

$N = 601$ $\qquad\qquad$ $N = 6{,}366$