ECO475H1 S
Applied Econometrics II

Tianyu Du

February 7, 2019

`Github Page https://github.com/TianyuDu/Spikey_UofT_Notes`
`Note Page TianyuDu.com/notes`

Contents

1 Lecture 3. Jan. 24 2019 . . . 2
1.1 Two Side Censoring MLE . . . 2
1.2 Two Side Truncated MLE . . . 3

2 Lecture 4. Jan. 31 2019 . . . 4
2.1 Tobit and Sample Selection . . . 4
2.2 Heckman Estimation (Two-Step Procedure) . . . 5

1 Lecture 3. Jan. 24 2019

1.1 Two Side Censoring MLE

Consider the latent dependent variable

$$Y^* = \mathbf{x}'\boldsymbol{\beta} + \epsilon \tag{1.1}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.
Therefore, given fixed $\mathbf{x}$,

$$Y^* \sim \mathcal{N}(\mathbf{x}'\boldsymbol{\beta}, \sigma^2) \tag{1.2}$$

Define parameter set

$$\boldsymbol{\theta} \equiv (\boldsymbol{\beta}, \sigma) \tag{1.3}$$

The observable variable is

$$Y = \begin{cases} U & \text{if } Y^* \geq U \\ Y^* & \text{if } Y^* \in (L, U) \\ L & \text{if } Y^* \leq L \end{cases} \tag{1.4}$$

Let $f_Y(y|\mathbf{x}, \boldsymbol{\beta}) : [L, U] \to [0, 1]$ be the probability measure of Y.
Let $y \in [L, U]$,

$$f_Y(y|\mathbf{x}, \boldsymbol{\beta}) = \begin{cases} \mathbb{P}(Y^* \geq U|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \geq U \\ f_{Y^*}(y|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \in (L, U) \\ \mathbb{P}(Y^* \leq L|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \leq L \end{cases} \tag{1.5}$$

$$= \begin{cases} 1 - F_{Y^*}(U|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \geq U \\ f_{Y^*}(y|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \in (L, U) \\ F_{Y^*}(L|\mathbf{x}, \boldsymbol{\beta}) & \text{if } y \leq L \end{cases} \tag{1.6}$$

Define indicator $(d_1(y), d_2(y), d_3(y))$ as

$$d_1(y) \equiv \mathcal{I}(y \geq U) \tag{1.7}$$

$$d_2(y) \equiv \mathcal{I}(y \in (L, U)) \tag{1.8}$$

$$d_3(y) \equiv \mathcal{I}(y \leq L) \tag{1.9}$$

Then the probability measure of Y can be expressed as

$$f_Y(y|\mathbf{x}, \boldsymbol{\beta}) = (1 - F_{Y^*}(U|\mathbf{x}, \boldsymbol{\beta}))^{d_1} \times f_{Y^*}(y|\mathbf{x}, \boldsymbol{\beta})^{d_2} \times F_{Y^*}(L|\mathbf{x}, \boldsymbol{\beta})^{d_3} \tag{1.10}$$

Suppose samples are i.i.d., the joint density is

$$f_{Y_1,\dots,Y_N}(y_1, \dots, y_N|\mathbf{X}, \boldsymbol{\beta}) = \prod_{i=1}^{N} f_Y(y_i|\mathbf{x}_i, \boldsymbol{\beta}) \tag{1.11}$$

The log-likelihood is

$$\mathcal{L}_N(\boldsymbol{\theta}|\mathbf{X}) = \sum_{i=1}^{N} \Big\{ d_{1,i} \times \ln(1 - F_{Y^*}(U|\mathbf{x}_i, \boldsymbol{\beta})) + d_{2,i} \times \ln(f_{Y^*}(y|\mathbf{x}_i, \boldsymbol{\beta})) + d_{3,i} \times \ln(F_{Y^*}(L|\mathbf{x}_i, \boldsymbol{\beta})) \Big\} \tag{1.12}$$

Finally, solving

$$\hat{\boldsymbol{\theta}}_{MLE} = (\hat{\boldsymbol{\beta}}_{MLE}, \hat{\sigma}_{MLE}) = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}}\, \mathcal{L}_N(\boldsymbol{\theta}) \tag{1.13}$$

1.2 Two Side Truncated MLE

Suppose the observations are truncated with lower and upper bounds L and U. Let the latent dependent variable be

$$Y^* = \mathbf{x}'\boldsymbol{\beta} + \epsilon \tag{1.14}$$

and

$$\epsilon \sim \mathcal{N}(0, \sigma^2) \tag{1.15}$$

which implies, for given $\mathbf{x}$,

$$Y^* \sim \mathcal{N}(\mathbf{x}'\boldsymbol{\beta}, \sigma^2) \tag{1.16}$$

Define parameter set

$$\boldsymbol{\theta} \equiv \{\boldsymbol{\beta}, \sigma\} \tag{1.17}$$

Observable random variable Y is

$$Y = \begin{cases} Y^* & \text{if } Y^* \in (L, U) \\ -- & \text{if } Y^* \notin (L, U) \end{cases} \tag{1.18}$$

Constructing the distribution for Y, note that F_Y is only defined on $y \in (L, U)$,

$$F_Y(y|\mathbf{x}, \boldsymbol{\theta}) = \mathbb{P}(Y < y|\mathbf{x}, \boldsymbol{\theta}) \tag{1.19}$$

$$= \frac{\mathbb{P}(Y^* < y \wedge Y^* \in (L, U)|\mathbf{x}, \boldsymbol{\theta})}{\mathbb{P}(Y^* \in (L, U)|\mathbf{x}, \boldsymbol{\theta})} \tag{1.20}$$

$$= \frac{\mathbb{P}(Y^* \in (L, y)|\mathbf{x}, \boldsymbol{\theta})}{\mathbb{P}(Y^* \in (L, U)|\mathbf{x}, \boldsymbol{\theta})} \tag{1.21}$$

$$= \frac{F_{Y^*}(y|\mathbf{x}, \boldsymbol{\theta}) - F_{Y^*}(L|\mathbf{x}, \boldsymbol{\theta})}{F_{Y^*}(U|\mathbf{x}, \boldsymbol{\theta}) - F_{Y^*}(L|\mathbf{x}, \boldsymbol{\theta})} \tag{1.22}$$

Then construct the density of Y

$$f_Y(y|\mathbf{x}, \boldsymbol{\theta}) = \frac{\partial F_Y(y|\mathbf{x}, \boldsymbol{\theta})}{\partial y} \tag{1.23}$$

$$= \frac{f_{Y^*}(y|\mathbf{x}, \boldsymbol{\theta})}{F_{Y^*}(U|\mathbf{x}, \boldsymbol{\theta}) - F_{Y^*}(L|\mathbf{x}, \boldsymbol{\theta})} \tag{1.24}$$

The sample log-likelihood is

$$\mathcal{L}_N(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln(f_{Y^*}(y_i|\mathbf{x}_i, \theta)) - \ln(F_{Y^*}(U|\mathbf{x}_i, \boldsymbol{\theta}) - F_{Y^*}(L|\mathbf{x}_i, \boldsymbol{\theta})) \tag{1.25}$$

and the estimator is given by

$$\hat{\boldsymbol{\theta}}_{MLE} = \{\hat{\boldsymbol{\beta}}_{MLE}, \hat{\sigma}_{MLE}\} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \, \mathcal{L}_N(\boldsymbol{\theta}) \tag{1.26}$$

2 Lecture 4. Jan. 31 2019

2.1 Tobit and Sample Selection

Model the *observable* variables in Tobit model with sample selection are determined by both **outcome equation** and **selection equation**.

$$y_i = \begin{cases} \mathbf{x}_i'\beta + \epsilon_i & \text{if } \mathbf{w}_i'\gamma + v_i > 0 \\ ✗ & \text{otherwise} \end{cases} \tag{2.1}$$

where **unmeasurable errors** are assumed to follow joint normal distribution,

$$\begin{pmatrix} \epsilon_i \\ v_i \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \begin{pmatrix} \sigma_\epsilon^2 & \rho\sigma^2 \\ \rho\sigma^2 & 1 \end{pmatrix}) \tag{2.2}$$

Lemma 2.1. If (ϵ, v) follows joint normal distribution, then there exists $e \perp v$ and $e \sim \mathcal{N}(0,1)$ such that

$$\frac{\epsilon}{\sigma_\epsilon} = \rho v + e \tag{2.3}$$

Expectation Define $\tilde{\mathbf{x}}_i \equiv [\mathbf{x}_i, \mathbf{w}_i]$, then the expected *observed* dependent variable is [1]

$$\mathbb{E}[y|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{2.4}$$
$$= \mathbb{E}[\mathbf{x}'\beta + \epsilon|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{2.5}$$
$$= \mathbf{x}'\beta + \mathbb{E}[\epsilon|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{2.6}$$
$$= \mathbf{x}'\beta + \mathbb{E}[\rho v\sigma_\epsilon + e\sigma_\epsilon|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{2.7}$$
$$= \mathbf{x}'\beta + \rho\sigma_\epsilon\mathbb{E}[v|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] + \sigma_\epsilon\mathbb{E}[e|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{2.8}$$
$$= \mathbf{x}'\beta + \rho\sigma_\epsilon\mathbb{E}[v|\mathbf{w}_i'\gamma + v_i > 0, \tilde{\mathbf{x}}] \tag{2.9}$$

Remark 2.1. If $\rho = 0$ in equation (2.9), there is no sample selection problem and we can use OLS to estimate the outcome equation.

Lemma 2.2. If $X \sim \mathcal{N}(\mu, \sigma^2)$ then

$$\mathbb{E}[X|X > \alpha] = \mu + \sigma\frac{\phi(\frac{x-\mu}{\sigma})}{1 - \Phi(\frac{x-\mu}{\sigma})} \tag{2.10}$$

(continue)

$$\cdots = \mathbf{x}'\beta + \rho\sigma_\epsilon\mathbb{E}[v|v > -\mathbf{w}'\gamma, \tilde{\mathbf{x}}] \tag{2.11}$$
$$= \mathbf{x}'\beta + \rho\sigma_\epsilon\frac{\phi(-\mathbf{w}'\gamma)}{1 - \Phi(-\mathbf{w}'\gamma)} \tag{2.12}$$
$$= \mathbf{x}'\beta + \rho\sigma_\epsilon\frac{\phi(\mathbf{w}'\gamma)}{\Phi(\mathbf{w}'\gamma)} \tag{2.13}$$
$$= \mathbf{x}'\beta + \rho\sigma_\epsilon\lambda(\mathbf{w}'\gamma) \tag{2.14}$$

where $\lambda(x)$ is the **inverse Mill's ratio** of standard normal at x.

[1] For each variable, the i subscript is omitted in the derivation

Marginal Effect Consider the case

$$\exists x_k \in \mathbf{x} \cap \mathbf{w} \tag{2.15}$$

for instance, x_k can be *wage taxation.* The marginal effect of x_k is

$$\begin{aligned}
\frac{\partial \mathbb{E}[y|\mathbf{w}'\gamma + v > 0, \tilde{\mathbf{x}}]}{\partial x_k} &= \frac{\partial \mathbf{x}'\beta + \rho\sigma_\epsilon \lambda(\mathbf{w}'\gamma)}{\partial x_k} && (2.16)\\
&= \beta_k + \rho\sigma_\epsilon \lambda'(\mathbf{w}'\gamma)\gamma_k && (2.17)\\
& && (2.18)
\end{aligned}$$

where β_k measures the **direct effect** and $\lambda'(\mathbf{w}'\gamma)\gamma_k$ measures the **indirect effect** of x_k.

2.2 Heckman Estimation (Two-Step Procedure)

Step 1 Run a *probit* estimation on the selection equation.
MLE gives

(i) An estimation $\hat{\gamma}_{MLE}$ captures the *indirect effect* of regressors in $\mathbf{w}$ on y through the selection equation.

And compute

$$\hat{\lambda}(\mathbf{w}'\hat{\gamma}_{MLE}) \equiv \frac{\phi(\mathbf{w}'\hat{\gamma}_{MLE})}{\Phi(\mathbf{w}'\hat{\gamma}_{MLE})} \tag{2.19}$$

Step 2 Run OLS

$$y = \mathbf{x}'\beta + \rho\sigma_\epsilon \hat{\lambda} + \eta \text{ where } \mathbb{E}[\eta|\mathbf{x}, \hat{\lambda}] = 0 \tag{2.20}$$

OLS gives

(i) An estimation $\hat{\beta}_{OLS}$ measures the *direct effect* of regressors in $\mathbf{x}$ on y through the outcome equation.

(ii) An estimation of $\widehat{\rho\sigma_\epsilon}$, given $\sigma_\epsilon > 0$, we can estimate the *sign* of ρ.

Special Case (i) Consider the special case where

$$\begin{aligned}
\mathbf{w} &= \mathbf{x} && (2.21)\\
\lambda(x) &\text{ is linear} && (2.22)
\end{aligned}$$

then (2.14) and regression (2.20) can be written as

$$\begin{aligned}
y &= \mathbf{x}'\beta + \rho\sigma_\epsilon \mathbf{x}'\lambda(\gamma) + \eta && (2.23)\\
&= \mathbf{x}'[\beta + \rho\sigma_\epsilon\lambda(\gamma)] + \eta && (2.24)
\end{aligned}$$

where $\beta + \rho\sigma_\epsilon\lambda(\gamma)$ represents the *mixed and non-separable* effect.

Special Case (ii) If

$$\mathbf{w} = [\mathbf{x}, z] \tag{2.25}$$
$$\lambda(x) \text{ is linear} \tag{2.26}$$
$$\tag{2.27}$$

Let the coefficients of $\mathbf{w}$ be $[\gamma, \theta]$, then

$$\lambda(\mathbf{w}[\gamma, \theta]) = \lambda(\mathbf{x}\gamma) + \lambda(z\theta) \tag{2.28}$$
$$= \mathbf{x}\lambda(\gamma) + z\lambda(\theta) \tag{2.29}$$

Then the regression can be rewritten as

$$y = \mathbf{x}'[\beta + \rho\sigma_\epsilon\lambda(\gamma)] + \rho\sigma_\epsilon z\lambda(\theta) + \eta \tag{2.30}$$

Remark 2.2. Therefore, if λ is linear, we need at least one exclusion variable to identify the direct and indirect effects. If λ is non-linear, it's *probably* fine.