Wooldridge chs 15-17:

Tobit Models, Bootstrapping

Type I Tobit

$$y^* = x'\beta + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2$$

$$y = \begin{cases} y^* & \text{if } y^* > 0 \\ 0 & \text{if } y^* \leq 0 \end{cases}$$

- MLE, NLS

Nothing really changes if we have a different censoring point

Generalized tobit type 1

$$y^* = x'\beta + \varepsilon$$

$$y = \begin{cases} y^* & y^* > y_{io} \\ 0 & y^* \leq y_{io} \end{cases} \qquad \varepsilon \sim N(0, \sigma^2)$$

y_{io} known but different.

Not robust to heteroskedasticity or nonnormality of ε.

Type II Tobit - Sample selection

- we observe y_1 if y_2^* exceeds some threshold

$y_1 = x_1'\beta_1 + \varepsilon_1$

$y_2 = 1\{x'\delta_2 + \varepsilon_2 > 0\}$

1] We observe (x, y_2)

2] We observe y_1 iff $y_2 = 1$

3] $\varepsilon_2 \sim N(0, \underbrace{1}_{\text{normalization}})$

4] $E[\varepsilon_1 | \varepsilon_2] = \eta_1 \varepsilon_2$. This will lead to inconsistency of OLS.

Generalization of type I Tobit. (ie if $y_1 = y_2$)

Heckman 2-step:

$$E[y_1 | x, y_2] = x_1'\beta_1 + \eta_1 E[\varepsilon_2 | x, y_2]$$

But we only have data when $y_2 = 1$, so

$$E[y_1 | x, 1] = x_1'\beta_1 + \eta_1 \lambda(x'\delta_2)$$

where $\lambda(t) = \dfrac{\varphi(t)}{\Phi(t)}$

Step 2: Regress $y_1 = x_1'\beta_1 + \eta_1 \hat{\lambda} + u_1$

Step 1: Obtain $\hat{\delta}_2 \Rightarrow \hat{\lambda} = (x'\hat{\delta}_2)$

Consistent for β_1. The std errors need some corrections appeal of 2-step method:

1] "Simple" to implement

2] Widely applicable

3] Less restrictive than conditional MLE approach

Weaknesses:

$$E[y_1|x,1] = x_1'\beta_1 + \gamma_1 \lambda(x'\delta)$$

$$= x_1'\beta_1 + \gamma_1 (\xi_1' x_1 + \xi_2' x_2)$$

- cannot identify β_1 if λ is linear

Semiparametric methods can help get around this. (Requires exclusion restrictions.)

If assume normality of $\binom{\varepsilon_1}{\varepsilon_2}$, can use MLE.

Type III Tobit

- selection equation is a type I Tobit
  - don't just get {Work, not work}. Get {Wage, not work}

$$y_1 = x_1'\beta_1 + \varepsilon_1$$

$$y_2 = \max\{x'\delta_2 + \varepsilon_2, 0\}$$

1] always observe (x, y_2)

2] Observe y_1 iff $y_2 > 0$

3] $\varepsilon_2 \sim N(0, \sigma_2^2)$

4] $E[\varepsilon_1|\varepsilon_2] = \gamma_1 \varepsilon_2$, so we have correlation

Step 1) Estimate selection eqn

Step 2) Plug in estimated coefficients and then run OLS

Type IV Tobit – ?

Type V Tobit – Switching regression

$$y = \begin{cases} y_1^* = x_1'\beta_1 + \varepsilon_1 & \text{if } y_3 = 1 \\ y_2^* = x_2'\beta_2 + \varepsilon_1 & \text{if } y_3 = 0 \end{cases}$$

$$y_3 = 1\{x_3'\beta_3 + \varepsilon_3 > 0\}$$

Bootstrapping

$$y = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

$$\begin{bmatrix} X_{2i} \\ X_{3i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}, \begin{bmatrix} (0.1)^2 & 0.005 \\ 0.005 & (0.1)^2 \end{bmatrix} \right) \qquad \beta = \begin{bmatrix} -2 \\ 2 \\ 2 \end{bmatrix}$$

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

- How to generate correlated normals?

Parametric bootstrap

For each b, we draw

$$\hat{y}_i^b = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \varepsilon_i, \quad \varepsilon_i \sim N(0, s^2)$$

where $\hat{\beta}, s^2$ are from original regression. Then regress $\hat{y}_i^b$ on $[1, X_{2i}, X_{3i}]$ to get $\hat{\beta}^b$. Repeat.

Residual bootstrap:

- Run OLS, get $\hat{\varepsilon}_i$
- For each b, randomly assign these residuals to the x's to get y.
  - $\tilde{y}_i^b = \hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{\varepsilon}_j$
- Then regress $\tilde{y}_i^b$ on $[1 \; x_{2i} \; x_{3i}]$ to get $\hat{\beta}^b$.
- Repeat.

JEP article: Brownstone and Valletta, 2001