AME

Week 12: Tobit

Sophie Bindslev, December 2022

UNIVERSITY OF COPENHAGEN

Today's Plan

- Censored Outcomes
- Tobit Likelihood
- Features of Interest
- Marginal Effects
- Your time to shine!

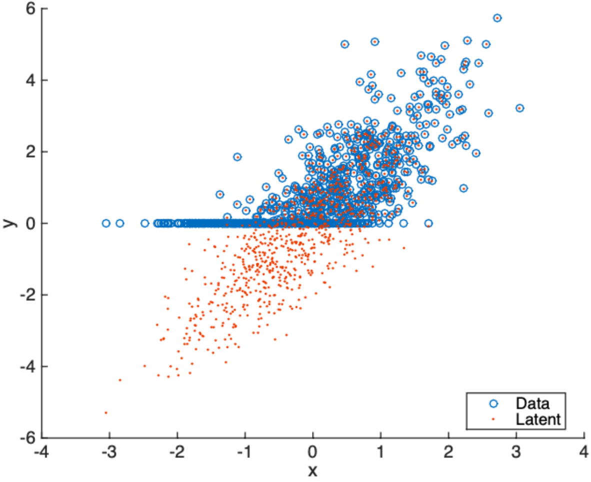
Censored outcomes I

- Sometimes we observe continuous variables but only above or below some threshold
- This may be due to "true" economic behaviour e.g. charitable donations (cannot be negative) or be due to data censoring e.g. top wealth censored in surveys
- OLS would yield inconsistent results even if the underlying (latent) outcome variable is appropriately described by a linear model
- Solution: Tobit Model

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2) \tag{1}$$

$$y_i = \max(y_i^*, 0) \tag{2}$$

Censored outcomes II



Deriving the Likelihood Function: Elephant Approach

- Tobit density:

$$f(y_i \mid \mathbf{x}_i) = \begin{cases} 0 & \text{if} \quad y_i < 0 \\ 1 - \Phi(\mathbf{x}_i\boldsymbol{\beta}/\sigma) & \text{if} \quad y_i = 0 \\ \frac{1}{\sigma}\phi(\mathbf{x}_i\boldsymbol{\beta}/\sigma) & \text{if} \quad y_i > 0 \end{cases} \tag{3}$$

  Anders derives this at: https://youtu.be/IwsE8Rr6l6E

- Taken together we get the log-likelihood contribution:

$$\begin{aligned} \ell_i(\boldsymbol{\theta}) =& \mathbf{1}\{y_i = 0\} \log \left[1 - \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)\right] \\ &+ \mathbf{1}\{y_i > 0\} \log \left[\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)\right] \end{aligned} \tag{4}$$

- Coding tip: $\mathbf{1}\{y_i = 0\}$ can be implemented using $(\mathtt{y} == 0)$

Features of interest

- We can obtain the conditional mean of the outcome variable in its censored and uncensored ranges:

$$\mathbb{E}(y_i|\mathbf{x}_i) = \mathbf{x}_i\beta\Phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right) + \sigma\phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right), \tag{5}$$

$$\mathbb{E}(y_i|\mathbf{x}_i, y_i > 0) = \mathbf{x}_i\beta + \sigma\lambda\left(\frac{\mathbf{x}_i\beta}{\sigma}\right). \tag{6}$$

  where $\lambda(.) = \phi(.)/\Phi(.)$ is the inverse mills ratio
- What would these formulas look like if there were no censoring but the data generating process were otherwise unchanged?
- Can you see how this links to inconsistent OLS estimates when there is censoring?

Marginal Effects

- The marginal effect of $\mathbf{x}_i$ conditional on an uncensored outcome variable:

$$\frac{\partial \mathrm{E}\left(y_i|\mathbf{x}_i, y_i > 0\right)}{\partial \mathbf{x}_i} = \boldsymbol{\beta}\Big\{1 - \lambda(\cdot)\big[\mathbf{x}_i\boldsymbol{\beta}/\sigma + \lambda(\cdot)\big]\Big\} \tag{7}$$

- The marginal effect of $\mathbf{x}_i$ on the mean of the outcome variable:

$$\frac{\partial \mathrm{E}\left(y_i|\mathbf{x}_i\right)}{\partial \mathbf{x}_i} = \boldsymbol{\beta}\Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) \tag{8}$$

Your time to shine!

- Fill in `tobit_ante.py` and solve the problem set
- When simulating data in Q1 remember that we're simulating data according to the censored outcome DGP, see eqs. (1) and (2)
- For plotting you can use `plt.hist(x, bins = 50)`, `plt.scatter()`, `plt.show()` if using np.ndarrays and `df.hist()` if using dataframes