Tobit Regression with Sample Selection

Model specification

March 9, 2019

Assume that the two real random variables Y_S^*, Y_O^* satisfy

$$\begin{aligned}
Y_S^* &= \beta_S' X_S + \varepsilon_S, &(1)\\
Y_O^* &= \beta_O' X_O + \varepsilon_O, \quad \text{where} &(2)\\
\begin{bmatrix}\varepsilon_S\\ \varepsilon_O\end{bmatrix} &\overset{|X_{S,O}}{\sim} \mathcal{N}_2\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho\sigma\\ \rho\sigma & \sigma^2\end{bmatrix}\right). &(3)
\end{aligned}$$

For $j = 1, \ldots, N$, we observe realizations of $\left(Y_{Sj}, Y_{Oj}, X_{Sj}', X_{Oj}'\right)$, where

$$Y_S \quad \begin{cases} 0 & Y_S^* \leq 0, \\ 1 & \text{else}, \end{cases} \qquad (4)$$

$$Y_O \quad \begin{cases} \text{NA} & Y_S = 0, \\ 0 & Y_S = 1, \text{ and } Y_O^* \leq T, \\ Y_O^* & Y_S = 1, \text{ and } Y_O^* > T, \end{cases} \qquad (5)$$

and where $0 \leq T < \infty$ is a fixed known threshold. We want to estimate the unknown parameters $\theta = \left(\beta_S', \beta_O', \rho, \sigma\right)$ using maximum likelihood which is (for one observation)

$$\begin{aligned}
L(\theta|y_o, y_s, T) &= \mathbb{P}\left(Y_O = \text{NA}\right)^{1-y_S} \mathbb{P}\left(Y_O = 0, Y_S = 1\right)^{y_S I\{y_O^* \leq T\}} f_{Y_O,Y_S}(\theta, y_O, y_S)^{y_S I\{y_O^* > T\}}, \\
&= \left(1 - \Phi\left(\beta_S' x_S\right)\right)^{1-y_S} \Phi_2\left(\frac{T - \beta_O' x_O}{\sigma}, \beta_S' x_S, -\rho\right)^{y_S I\{y_O^* > T\}} * \ldots \\
&\ldots * \left(\frac{1}{\sigma}\phi\left(\frac{y_O - \beta_O' x_O}{\sigma}\right) \Phi\left(\frac{\rho/\sigma\left(y_O - \beta_O' x_O\right) + \beta_S' x_S}{\sqrt{1-\rho^2}}\right)\right)^{y_S I\{y_O^* > T\}},
\end{aligned}$$

where f is the joint density of Y_O, Y_S, and $\phi, \Phi,$ and Φ_2 are the pdf, cdf and joint cdf of the standard normal random variable(s with correlation coefficient ρ).