TA session #6
ECON 342

J. Marcelo Ochoa

February 25, 2009

Problem 1. Assume that you have a doubly censored variable, such that:

$$y_i^* = x_i\beta + \varepsilon_i$$

$$y_i = \begin{cases} a_1, & \text{if } y_i^* \leq a_1 \\ y_i^*, & \text{if } a_1 < y_i^* < a_2 \\ a_2, & \text{if } y_i^* \geq a_2 \end{cases}$$

(a) Find $\mathbb{P}(y_i = a_1|x_i)$, $\mathbb{P}(y_i = a_2|x_i)$ and $\mathbb{P}(y_i \leq z|x_i)$ for $z \in (a_1, a_2)$.

(b) Find $\mathbb{E}(y_i|x_i, a_1 < y_i < a_2)$ and $\mathbb{E}(y_i|x_i)$.

(c) Consider the following method for estimating β. Using only the uncensored observations, run the OLS regression of y_i on x_i, Does this method produce a consistent estimator of β?

(d) What is the partial effect of x_{ij}?

(e) For data censoring, how would the analysis change if a_1 and a_2 were replaced with a_{i1} and a_{i2}?

Solution. Using the definition of y_i we have,

$$\begin{aligned} \mathbb{P}(y_i = a_1|x_i) =& \mathbb{P}(y_i^* \leq a_1|x_i) \\ =& \mathbb{P}(\varepsilon_i \leq a_1 - x_i\beta|x_i) \\ =& \mathbb{P}\left(\frac{\varepsilon_i}{\sigma} \leq \frac{a_1 - x_i\beta}{\sigma}|x_i\right) \\ =& \Phi\left(\frac{a_1 - x_i\beta}{\sigma}\right) \end{aligned}$$

and,

$$\begin{aligned}\mathbb{P}\left(y_i = a_2|x_i\right) =& \mathbb{P}\left(y_i^* \geq a_2|x_i\right)\\ =& \mathbb{P}\left(\varepsilon_i \geq a_2 - x_i\beta|x_i\right)\\ =& \mathbb{P}\left(\frac{\varepsilon_i}{\sigma} \geq \frac{a_2 - x_i\beta}{\sigma}|x_i\right)\\ =& 1 - \Phi\left(\frac{a_2 - x_i\beta}{\sigma}\right)\end{aligned}$$

In the third case, note that when $y_i = y_i^*$ for values of $y_i^* \in (a_1, a_2)$, then,

$$\mathbb{P}\left(y_i \leq z|x_i\right) = \mathbb{P}\left(y_i^* \leq z|x_i\right) = \Phi\left(\frac{z - x_i\beta}{\sigma}\right)$$

To obtain the conditional expectation $\mathbb{E}\left(y_i|x_i, a_1 < y_i < a_2\right)$ from the equation above note,

$$f_y(z|x) = \mathbb{P}\left(y_i^* = z|x_i\right) = \frac{\partial\mathbb{P}\left(y_i^* \leq z|x_i\right)}{\partial z} = \frac{1}{\sigma}\phi\left(\frac{z - x_i\beta}{\sigma}\right)$$

then,

$$f_y(z|x, a_1 < z < a_2) = \mathbb{P}\left(y_i^* = z|x_i, a_1 < z < a_2\right) = \frac{f(z|x)}{F(a_2|x) - F(a_1|x)}$$

which using the above definitions equals,

$$f_y(z|x, a_1 < z < a_2) = \frac{\frac{1}{\sigma}\phi\left(\frac{z - x_i\beta}{\sigma}\right)}{\Phi\left(\frac{a_2 - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a_1 - x_i\beta}{\sigma}\right)}$$

Then,

$$\mathbb{E}\left(y_i|x_i, a_1 < y_i < a_2\right) = \mathbb{E}\left(y_i^*|x_i, a_1 < y_i^* < a_2\right) = \int_{a_1}^{a_2} y_i^* \frac{\frac{1}{\sigma}\phi\left(\frac{y_i^* - x_i\beta}{\sigma}\right)}{\Phi\left(\frac{a_2 - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a_1 - x_i\beta}{\sigma}\right)} dy_i^* \quad (1)$$

To solve the integral, re-write (1) as,

$$\mathbb{E}\left(y_i^*|x_i, a_1 < y_i^* < a_2\right) = \frac{1}{\Phi\left(\frac{a_2 - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a_1 - x_i\beta}{\sigma}\right)} \int_{a_1}^{a_2} \frac{y_i^* - x_i\beta}{\sigma}\phi\left(\frac{y_i^* - x_i\beta}{\sigma}\right) + \frac{x_i\beta}{\sigma}\phi\left(\frac{y_i^* - x_i\beta}{\sigma}\right) dy_i^*$$

and let $u = \frac{y_i^* - x_i\beta}{\sigma}$ thus $\sigma du = dy_i^*$, then:

$$\sigma\int_{\frac{a_1 - x_i\beta}{\sigma}}^{\frac{a_2 - x_i\beta}{\sigma}} u\phi(u)du = \sigma\phi\left(\frac{a_1 - x_i\beta}{\sigma}\right) - \sigma\phi\left(\frac{a_2 - x_i\beta}{\sigma}\right) \quad (2)$$

which follows by noting that $d\phi(u)/du = -u\phi(u)$. Similarly,

$$\frac{x_i\beta}{\sigma}\int_{a_1}^{a_2} \phi\left(\frac{y_i^* - x_i\beta}{\sigma}\right) dy_i^* = \frac{x_i\beta}{\sigma}\sigma\int_{\frac{a_1 - x_i\beta}{\sigma}}^{\frac{a_2 - x_i\beta}{\sigma}} \phi(u)du = x_i\beta\left[\Phi\left(\frac{a_2 - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a_1 - x_i\beta}{\sigma}\right)\right] \quad (3)$$

With (2) and (3) we obtain:

$$\begin{aligned}\mathbb{E}\left(y_i|x_i, a_1 < y_i < a_2\right) =& \frac{1}{\Phi\left(\frac{a_2 - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a_1 - x_i\beta}{\sigma}\right)}\left\{\sigma\phi\left(\frac{a_1 - x_i\beta}{\sigma}\right) - \sigma\phi\left(\frac{a_2 - x_i\beta}{\sigma}\right)\right. \\ &\left. + x_i\beta\left[\Phi\left(\frac{a_2 - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a_1 - x_i\beta}{\sigma}\right)\right]\right\} \\ =& x_i\beta + \sigma\frac{\phi\left(\frac{a_1 - x_i\beta}{\sigma}\right) - \phi\left(\frac{a_2 - x_i\beta}{\sigma}\right)}{\Phi\left(\frac{a_2 - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a_1 - x_i\beta}{\sigma}\right)}\end{aligned}$$

Then, we can calculate $\mathbb{E}\left(y_i|x_i\right)$ as follows

$$\begin{aligned}\mathbb{E}\left(y_i|x_i\right) =& a_1\mathbb{P}\left(y_i = a_1|x_i\right) + a_2\mathbb{P}\left(y_i = a_2|x_i\right) + \mathbb{E}\left(y_i|x_i, a_1 < y_i < a_2\right)\mathbb{P}\left(a_1 < y_i < a_2|x_i\right) \\ =& a_1\Phi\left(\frac{a_1 - x_i\beta}{\sigma}\right) + a_2\left[1 - \Phi\left(\frac{a_2 - x_i\beta}{\sigma}\right)\right] + \left[\Phi\left(\frac{a_2 - x_i\beta}{\sigma}\right) - \Phi\left(\frac{a_1 - x_i\beta}{\sigma}\right)\right]x_i\beta \\ &+ \sigma\left[\phi\left(\frac{a_1 - x_i\beta}{\sigma}\right) - \phi\left(\frac{a_2 - x_i\beta}{\sigma}\right)\right] \\ \neq& x_i\beta\end{aligned}$$

which also suggests that estimating β using OLS will not produce a consistent estimate. Similarly, we found that

$$\mathbb{E}\left(y_i|x_i, a_1 < y_i < a_2\right) \neq x_i\beta$$

which also suggests that running a regression using only uncensored observations will lead to an inconsistent estimate of β.

$\Diamond$