Answer key to problem set # 4

ECON 342

Prof. Shakeeb Khan

T.A. J. Marcelo Ochoa

Spring, 2009

Problem (3)**.** Let $y_i = max(x_i'\beta_0 + \epsilon_i, 0)$. Formally derive the expression for $\mathbb{E}(y_i|x_i, y_i > 0)$ and $\mathbb{E}(y_i|x_i)$ assuming ϵ_i is normally distributed with variance σ_0^2.

Solution. Let $y_i^* = x_i'\beta_0 + \epsilon_i$, then it must be that $y_i^* \sim \mathcal{N}(x_i'\beta_0, \sigma_0^2)$. Using this latent variable approach the model proposed can be re-written as,

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

When $y = 0$ we have

$$\Pr(y = 0) = Pr(y^* \leq 0) = \Phi\left(\frac{-x_i'\beta_0}{\sigma}\right) = 1 - \Phi\left(\frac{x_i'\beta_0}{\sigma}\right).$$

Then,

$$\mathbb{E}(y_i|x_i, y_i > 0) = \mathbb{E}(y_i^*|x_i, y_i^* > 0)$$

When $y_i = y_i^*$ the distribution that applies is the one of y_i^*. We know that $y_i^* \sim \mathcal{N}(x_i'\beta_0, \sigma_0^2)$ therefore the conditional probability can be written as,

$$f(y_i^*|y_i^* > 0) = \frac{(1/\sigma)\phi\left(\frac{y_i^* - x_i'\beta_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-x_i'\beta_0}{\sigma_0}\right)}$$

Then,

$$\mathbb{E}(y_i^*|x_i, y_i^* > 0) = \int_0^\infty y_i^* \frac{(1/\sigma)\phi\left(\frac{y_i^* - x_i'\beta_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-x_i'\beta_0}{\sigma_0}\right)} dy_i^* \tag{1}$$

To solve the integral re-write () as,

$$\mathbb{E}(y_i^*|x_i, y_i^* > 0) = \int_0^\infty \frac{\frac{y_i^* - x_i'\beta_0}{\sigma_0}\phi\left(\frac{y_i^* - x_i'\beta_0}{\sigma_0}\right) + \frac{x_i'\beta_0}{\sigma_0}\phi\left(\frac{y_i^* - x_i'\beta_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{-x_i'\beta_0}{\sigma_0}\right)} dy_i^*$$

and let $x = \frac{y_i^* - x_i'\beta_0}{\sigma_0}$ thus $\sigma_0 dx = dy_i^*$, then we have that solutions to the first part and second part of the integral are:

$$
\begin{aligned}
(1) \Rightarrow & \frac{\sigma_0}{1-\Phi\left(\frac{-x_i'\beta_0}{\sigma_0}\right)} \int_{-\frac{x_i'\beta_0}{\sigma}}^{\infty} x\phi(x)dx \\
&= \frac{\sigma_0}{1-\Phi\left(\frac{-x_i'\beta_0}{\sigma_0}\right)} \int_{-\frac{x_i'\beta_0}{\sigma}}^{\infty} -\frac{d\phi(x)}{dx}dx \\
&= \frac{\sigma_0}{1-\Phi\left(\frac{-x_i'\beta_0}{\sigma_0}\right)} \left[-\phi(\infty) + \phi\left(-\frac{x_i'\beta_0}{\sigma}\right)\right] \\
&= \frac{\sigma_0\phi\left(\frac{-x_i'\beta_0}{\sigma}\right)}{1-\Phi\left(\frac{-x_i'\beta_0}{\sigma_0}\right)} \\
&= \frac{\sigma_0\phi\left(\frac{-x_i'\beta_0}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)} \\
(2) \Rightarrow & \frac{x_i'\beta_0}{1-\Phi\left(\frac{-x_i'\beta_0}{\sigma_0}\right)} \int_0^{\infty} \frac{1}{\sigma_0}\phi\left(\frac{y_i^* - x_i'\beta_0}{\sigma_0}\right)dy_i^* \\
&= \frac{x_i'\beta_0}{1-\Phi\left(\frac{-x_i'\beta_0}{\sigma_0}\right)} \left(1-\Phi\left(\frac{-x_i'\beta_0}{\sigma_0}\right)\right) \\
&= x_i'\beta_0
\end{aligned}
$$

Using these results we have,

$$
\mathbb{E}\left(y_i|x_i, y_i > 0\right) = x_i'\beta_0 + \frac{\sigma_0\phi\left(-\frac{x_i'\beta_0}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)} \tag{2}
$$

The unconditional mean of y_i is,

$$
\begin{aligned}
\mathbb{E}\left(y_i|x_i\right) &= \mathbb{E}\left(y_i|x_i, y_i > 0\right)\Pr(y_i > 0) + \mathbb{E}\left(y_i|x_i, y_i = 0\right)\Pr(y_i^* \leq 0) \\
&= \mathbb{E}\left(y_i|x_i, y_i > 0\right)\Pr(x_i'\beta_0 + \epsilon_i > 0) + 0 \cdot \Pr(y_i^* \leq 0) \\
&= \left[x_i'\beta_0 + \frac{\sigma_0\phi\left(\frac{-x_i'\beta_0}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)}\right]\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right) \\
&= x_i'\beta_0\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right) + \sigma_0\phi\left(-\frac{x_i'\beta_0}{\sigma}\right)
\end{aligned}
$$

$\Diamond$

Problem (4)**.** Now suppose we alter the basic censored regression model in the following way:

$$y_i^* = x_i'\beta_0 + \epsilon_i,\ \epsilon_i \sim \mathcal{N}(0, \sigma_0^2)$$
$$y_i = c\mathcal{I}(y_i^* \leq 0) + y_i^*\mathcal{I}(y_i^* > 0)$$

where c is some constant.

(a) Find $\mathbb{E}\left(y_i|x_i, c\right)$

(b) How does the MLE change with c?

Solution. From the last question we have:

$$\mathbb{E}\left(y_i|x_i, y_i > 0\right) = \mathbb{E}\left(y_i^*|x_i, y_i^* > 0\right) = x_i'\beta_0 + \frac{\sigma_0\phi\left(\frac{x_i'\beta_0}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)} \tag{3}$$

then, we can calculate the expected value as:

$$\begin{aligned}
\mathbb{E}\left(y_i|x_i, c\right) &= \mathbb{E}\left(y_i|x_i, y_i > 0\right)\Pr(y_i > 0) + \mathbb{E}\left(y_i|x_i, y_i = c\right)\Pr(y_i^* \leq 0) \\
&= x_i'\beta_0\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right) + \sigma_0\phi\left(\frac{x_i'\beta_0}{\sigma}\right) + c\Pr(\epsilon_i \leq -x_i'\beta_0) \\
&= x_i'\beta_0\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right) + \sigma_0\phi\left(\frac{x_i'\beta_0}{\sigma}\right) + c\Phi\left(\frac{-x_i'\beta_0}{\sigma_0}\right) \\
&= x_i'\beta_0\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right) + \sigma_0\phi\left(\frac{x_i'\beta_0}{\sigma}\right) + c\left[1 - \Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\right]
\end{aligned}$$

When $c < 0$ the MLE estimate does not change with c. Note that,

$$\Pr(y_i = c) = \Pr(y_i^* \leq 0) = 1 - \Phi\left(\frac{x_i'\beta_0}{\sigma}\right)$$

thus, the log-likelihood is given by,

$$\mathcal{L}(\beta, \sigma) = \sum_{y_i = c} \ln\left[1 - \Phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right)\right] + \sum_{y_i > 0} \ln\left[\frac{1}{\sqrt{2\pi\sigma}}\exp\left[-\frac{(Y_i - \mathbf{x}_i\beta)^2}{2\sigma}\right]\right]$$

while in case $c = 0$ we have the same result.

In the case when $c > 0$, we can have,

$$\Pr(y_i = c) = \Pr(y_i^* \leq 0) + \Pr(y_i^* = c) = 1 - \Phi\left(\frac{x_i'\beta_0}{\sigma}\right) + \frac{1}{\sigma}\phi\left(\frac{c - x_i'\beta_0}{\sigma}\right)$$

thus, the log-likelihood is given by,

$$\mathcal{L}(\beta, \sigma) = \sum_{y_i = c} \ln\left[1 - \Phi\left(\frac{\mathbf{x}_i\beta}{\sigma}\right) + \frac{1}{\sigma}\phi\left(\frac{c - x_i'\beta_0}{\sigma}\right)\right] + \sum_{y_i > 0, y_i \neq c} \ln\left[\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta_0}{\sigma}\right)\right]$$

It is evident, that in this case the MLE of β will be numerically different as it depends on c. ◊