ECON 203C: System Models
TA Note 8: Version 2

Tobit Models

Hisayuki Yoshimoto
Last Modified: June 6, 2008

Abstract: In section 2, we introduce the type I standard Tobit model. In section 3 and 4, we solve both of type I standard and top-coded Tobit models.

1 Tobit Models

Tobit model is the model in which econometricians only observe dependent variables that satisfy some restrictions. For example, if dependent variable y_i is household i's expenditure on durable goods (such as car and house) we observe

$$y_i = \begin{cases} \$ \text{ positive number} & : \text{if household } i \text{ purchases durable goods} \\ \$ \ 0 & : \text{if household } i \text{ does not purchase a durable good} \end{cases},$$

associate with dependent variable such as

$$x_i = \begin{bmatrix} 1 \\ \text{income of } i \\ \# \text{ of family member of } i \\ \text{suburb dummy} \end{bmatrix}.$$

With this type of restricted dependent variables, OLS provides a biased estimator (we will discuss the reason of bias later), and we have to employ alternative methods, Tobit models. According to Amemiya (Advanced Econometrics, 1985) there are five types of Tobit models, i.e. type I to V Tobit models. We will discuss type I standard Tobit model (section 2 and 3) and type I top-coded Tobit model (in section 4).

2 Type I Standard Tobit Model

2.1 Type I Standard Tobit Model

Type I (standard) Tobit model is described as follows. The latent variable y_i^* has a linear form

$$\underbrace{y_i^*}_{1\times 1} = \underbrace{x_i'}_{1\times K}\underbrace{\beta}_{K\times 1} + \underbrace{\varepsilon_i}_{1\times 1}$$

where

$$\varepsilon_i | \, x_i \sim N\left(0, \sigma^2\right).$$

The dependent variable y_i^* is determined by

$$y_i = \begin{cases} y_i^* & : \text{ if } y_i^* > L \\ 0 & : \text{ if } y_i^* \leq L \end{cases}. \tag{1}$$

Explaining examples to get intuitions. For example, y_i is a expenditure of durable goods[1] that takes positive values only if a household spend money on them. Another example is that y_i is working hours of married woman, who spend positive amount of time on working only if she decides to work.

[1] Your TA haven't purchased any durable goods (car, TV, computer, ect) for last two years. Obviously, his expendititure on durable goods (y_i) is zero, unfortunately.

Notice that if x_i includes a constant, we can normalize $L = 0$. Formally, assume

$$x_i = \begin{bmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{iK} \end{bmatrix} \quad \text{and } \beta = \begin{bmatrix} \beta_{const} \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix},$$

Then,

$$\begin{aligned} y_i^* &> L \\ &\Leftrightarrow x_i'\beta + \varepsilon_i > L \qquad (\text{since } y_i^* = x_i'\beta + \varepsilon_i) \\ &\Leftrightarrow \beta_{const} + \begin{bmatrix} x_{i1} \\ \vdots \\ x_{iK} \end{bmatrix}' \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} + \varepsilon_i > L \\ &\Leftrightarrow (\beta_{const} - L) + \begin{bmatrix} x_{i1} \\ \vdots \\ x_{iK} \end{bmatrix}' \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} + \varepsilon_i > 0. \end{aligned}$$

Thus, we cannot separately estimate β_{const} and L. Thus, we can normalize $L = 0$ without loss of generality. Let's draw the graph to get intuition.

It is obvious from the graph that OLS provides a biased estimator. In Tobit models, we have three goals,
Goal 1: consistently estimate β
Goal 2: derive conditional expectation $E[y_i | x_i, y_i > 0]$
Goal 3: derive conditional expectation $E[y_i | x_i]$.

Actually, goal 2 and 3 have policy implications such as "how much does the expenditure on durable goods increase as income increases?".

2.2 Maximum Likelihood Estimation

First of all, we derive the conditional density function of y_i. Define the f and f^* as

$$\begin{cases} f & : \text{ pdf of } y_i \\ f^* & : \text{ pdf of } y_i^* \end{cases}$$

Remind that y_i is determined by (rewriting definition of y_i ,(1), with $L = 0$)

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

Now, we derive the conditional distribution function of y_i, i.e. $f\left(y_i | x_i\right)$. According to the definition of y_i above we have to consider two cases, (i) case of $y_i > 0$ and (ii) case of $y_i = 0$.

(i) Case of $y_i > 0$

The above definition of y_i indicates that if $y_i > 0$ the conditional distribution of y_i is the same as that of y_i^*. Therefore, if $y > 0$, we have

$$\begin{aligned} f\left(y_i | x_i\right) &= f^*\left(y_i | x_i\right) \\ &= f^*\left(y_i^* | x_i\right) \qquad (\text{since } y_i = y_i^* \text{ if } y_i > 0) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{\left(y_i - x_i'\beta\right)^2}{\sigma^2}\right) \qquad (\text{since } y_i^* \text{ is distributed normally } y_i^* \sim N\left(x_i'\beta, \sigma^2\right)) \\ &= \underbrace{\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2\right)}_{=\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)} \\ &= \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right). \qquad (\text{since pdf of standard normal is } \phi(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right), \text{ now } z = \frac{y_i - x_i'\beta}{\sigma}) \end{aligned}$$

(ii) Case of $y_i = 0$

On the other hand, for $y_i = 0$, we have the mass conditional probability $\Pr\left(y_i = 0 | x_i\right)$ which is equal to

$$\begin{aligned} \Pr\left(y_i = 0 | x_i\right) &= \Pr\left(y_i^* < 0 | x_i\right) \\ &= \Pr\left(x_i'\beta + \varepsilon_i \leq 0 | x_i\right) \qquad (\text{by definition of latent variable } y_i^*,\ y_i^* = x_i'\beta + \varepsilon_i) \\ &= \Pr\left(\underbrace{\varepsilon_i}_{\text{distributed as } N(0,\sigma^2)} \leq -x_i'\beta \,\middle|\, x_i\right) \\ & \quad \Pr\left(\underbrace{\frac{\varepsilon_i}{\sigma}}_{\text{distributed as standard normal}} \leq -\frac{x_i'\beta}{\sigma} \,\middle|\, x_i\right) \\ &= \Phi\left(-\frac{x_i'\beta}{\sigma}\right) \qquad (\text{where } \Phi \text{ is the c.d.f. of standard normal}) \\ &= 1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right) \qquad (\text{since standard normal distribution is symmetric, } \Phi(-z) = 1 - \Phi(z)). \end{aligned}$$

Therefore, according to the result of (i) and (ii), the conditional density function is expressed as

$$f\left(y_i | x_i\right) = \begin{cases} \text{continuous part} & f^*\left(y_i | x_i\right) & = \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) & : \text{ if } y_i > 0 \\ \text{mass part} & \Pr\left(y_i = 0 | x_i\right) & = 1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right) & : \text{ if } y_i \leq 0 \end{cases} \tag{2}$$

Next, define the dummy variable

$$d_i = \begin{cases} 1 & : \text{ if } y_i > 0 \\ 0 & : \text{ if } y_i \leq 0 \end{cases} . \tag{3}$$

Then, by (2) and (3), we can express the conditional pdf of y_i given x_i as[2]

$$\begin{aligned} f\left(y_i|\, x_i\right) &= \left\{f^*\left(y_i|\, x_i\right)\right\}^{d_i} \cdot \left\{\Pr\left(y_i = 0|\, x_i\right)\right\}^{1-d_i} \\ &= \left\{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\}^{d_i} \cdot \left\{1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right\}^{1-d_i}. \end{aligned}$$

Then, the likelihood function is defined as

$$\begin{aligned} L_N\left(\beta, \sigma^2\right) &= \prod_{i=1}^{N} f\left(y_i|\, x_i\right) \qquad \text{(since samples are i.i.d.)} \\ &= \prod_{i=1}^{N} \left\{f\left(y_i|\, x_i, y_i > 0\right)\right\}^{d_i} \cdot \left\{\Pr\left(y_i^* \le 0|\, x_i\right)\right\}^{1-d_i} \\ &= \prod_{i=1}^{N} \left\{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\}^{d_i} \cdot \left\{1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right\}^{1-d_i}. \end{aligned}$$

The log-likelihood function is

$$\begin{aligned} l_n\left(\beta, \sigma^2\right) &= \ln L_N\left(\beta, \sigma^2\right) \\ &= \sum_{i=1}^{N} \left[d_i \ln\left\{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\} + (1 - d_i)\ln\left\{1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right\}\right]. \end{aligned}$$

The maximum likelihood estimator is define as

$$\left(\hat\beta_{ML}, \hat\sigma^2_{ML}\right) = \arg\max_{\beta, \sigma^2}\left\{l_n\left(\beta, \sigma^2\right)\right\}.$$

Asymptotic distribution of MLE is given by (as the same as other ML estimator asymptotic distribution)

$$\sqrt{N}\left(\left[\begin{array}{c} \hat\beta_{ML} \\ \hat\sigma^2_{ML} \end{array}\right] - \left[\begin{array}{c} \beta \\ \sigma^2 \end{array}\right]\right) \xrightarrow{d} N\left(0_{(K+1)\times 1}, I_1^{-1}\right),$$

where

$$I_1^{-1} = -E\left[\frac{\partial}{\partial\beta'\partial\beta} l_1\left(\beta, \sigma^2\right)\right].$$

Extra Note:
Deriving the f.o.c. equations of ML estimator. (you might skip this part, this is just mathematical calculation)
The log-likelihood function can be written as

$$\begin{aligned} l_n\left(\beta, \sigma^2\right) &= \sum_{i=1}^{N}\left[d_i \ln\left\{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\} + (1 - d_i)\ln\left\{1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right\}\right] \\ &= \sum_{i=1}^{N}\left[d_i \ln\left\{\frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2\right)\right\} + (1 - d_i)\ln\left\{1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right\}\right] \\ &= \sum_{i=1}^{N}\left[d_i\left\{-\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\ln(2\pi) - \frac{1}{2}\frac{\left(y_i - x_i'\beta\right)^2}{\sigma^2}\right\} + (1 - d_i)\ln\left\{1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right\}\right] \end{aligned}$$

Taking f.o.c. w.r.t. β

$$\begin{aligned} \underbrace{\frac{\partial}{\partial\beta} l_n\left(\beta, \sigma^2\right)}_{K\times 1} &= \sum_{i=1}^{N}\left[d_i\left\{-\frac{1}{2}\frac{y_i - x_i'\beta}{\sigma^2}x_i\right\} - (1 - d_i)\left\{\frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)}x_i\right\}\right] \\ &= \sum_{i=1}^{N}\left[d_i\left\{-\frac{1}{2}\frac{y_i - x_i'\beta}{\sigma^2}\right\} + (1 - d_i)\ln\left\{1 - \phi\left(\frac{x_i'\beta}{\sigma}\right)\right\}\right]\underbrace{x_i}_{K\times 1} = 0_{K\times 1} \end{aligned}$$

[2] It is sometimes convenient to write the equivalent indicator function notation

$$f\left(y_i|\, x_i\right) = \left\{f\left(y_i|\, x_i, y_i > 0\right)\right\}^{\mathbf{1}\{y_i > 0\}} \cdot \left\{\Pr\left(y_i^* \le 0|\, x_i\right)\right\}^{\mathbf{1}\{y_i \le 0\}} \qquad \text{(equivalent indicator function notation)}$$

Taking f.o.c. w.r.t. σ^2

$$\begin{aligned}
\underbrace{\frac{\partial}{\partial \sigma^2} l_n\left(\beta, \sigma^2\right)}_{1\times 1} &= \sum_{i=1}^{N}\left[d_i\left\{-\frac{1}{2}\frac{1}{\sigma^2}+\frac{1}{2}\frac{\left(y_i - x_i'\beta\right)^2}{\sigma^4}\right\}+\left(1-d_i\right)\frac{-\phi\left(\frac{x_i'\beta}{\sigma}\right)\cdot\left(-1\right)x_i'\beta\frac{1}{2\sigma^3}}{1-\Phi\left(\frac{x_i'\beta}{\sigma}\right)}\right] \\
&= \sum_{i=1}^{N}\left[d_i\left\{-\frac{1}{2}\frac{1}{\sigma^2}+\frac{1}{2}\frac{\left(y_i - x_i'\beta\right)^2}{\sigma^4}\right\}+\left(1-d_i\right)\frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{1-\Phi\left(\frac{x_i'\beta}{\sigma}\right)}\cdot x_i'\beta\frac{1}{2\sigma^3}\right] = 0_{1\times 1}.
\end{aligned}$$

Therefore, ML estimator is given by the solution of above f.o.c. system equations.

2.3 Conditional Expectations

In practice, we are interested in conditional expectations $E_{y_i|x_i,y_i>0}\left[y_i|\, x_i, y_i>0\right]$ and $E_{y_i|x_i}\left[y_i|\, x_i\right]$ to argue policy implications. The calculation of these conditional expectations requires some math skills. Now, we omit sample index i for simplicity of notation

$$\begin{aligned}
& E_{y|x}\left[y|\, x\right] \\
= \;& E_{\varepsilon|x}\left[y|\, x\right] \\
= \;& E_d\left[E_{\varepsilon|x,d}\left[y|\, x, d\right]\big|\, x\right] \qquad \text{(expectation trick, where } d \text{ is dummy defined in (3))} \\
= \;& \Pr\left(d=0|\right)\cdot E_{\varepsilon|x,d=0}\left[y|\, x, d=0\right] + \Pr\left(d=1|\, x\right)\cdot E_{\varepsilon|x,d=1}\left[y|\, x, d=1\right] \qquad \text{(definition of expectation)} \\
= \;& \Pr\left(y^*\le 0|\, x_i\right)\cdot \underbrace{E_{\varepsilon|x,d=0}\left[y|\, x, d=0\right]}_{=0 \text{ by definition of } d} + \Pr\left(y^* > 0|\, x\right)\cdot E_{\varepsilon|x,d=1}\left[y|\, x, d=1\right] \\
= \;& \Pr\left(y^* > 0|\, x_i\right)\cdot E_{\varepsilon|x,d=1}\left[y|\, x, d=1\right] \qquad \text{(since } y=0 \text{ if } d=0\text{)} \\
= \;& \Pr\left(y^* > 0|\, x_i\right)\cdot \underbrace{E_{\varepsilon|x,y>0}\left[y|\, x, y>0\right]}_{\text{calculating below}} \qquad \text{(since } y>0 \text{ if } d=1\text{)}
\end{aligned}$$

where

$$\begin{aligned}
E_{\varepsilon|x,y>0}\left[y|\, x, y>0\right] &= E_{\varepsilon|x,y^*>0}\left[y^*|\, x, y^*>0\right] \qquad \text{(since } y>0 \text{ is equivalent to } y^*>0\text{)} \\
&= E_{\varepsilon|x,x'\beta+\varepsilon>0}\left[x'\beta+\varepsilon|\, x, x'\beta+\varepsilon>0\right] \\
&= E_{\varepsilon|x'\beta+\varepsilon>0}\left[x'\beta+\varepsilon|\, x'\beta+\varepsilon>0\right] \qquad \text{(deleting redundant notation)} \\
&= E_{\varepsilon|\varepsilon>-x_i'\beta}\left[x_i'\beta+\varepsilon|\, \varepsilon>-x_i'\beta\right] \\
&= \underbrace{E_{\varepsilon|\varepsilon>-x_i'\beta}\left[x_i'\beta|\, \varepsilon>-x_i'\beta\right]}_{=x_i'\beta} + E_{\varepsilon|\varepsilon>-x_i'\beta}\left[\varepsilon|\, \varepsilon>-x_i'\beta\right] \qquad \text{(expectation is linear)} \\
&= x_i'\beta + \underbrace{E_{\varepsilon|\varepsilon>-x_i'\beta}\left[\varepsilon|\, \varepsilon>-x_i'\beta\right]}_{\text{calculating}}
\end{aligned}$$

Next, for calculating $E_{\varepsilon|\varepsilon>-x_i'\beta}\left[\varepsilon|\, \varepsilon>-x_i'\beta\right]$, we need to discuss the expectation of truncated standard normal distribution.

Assume that r.v. $z\sim N\left(0_{1\times 1}, 1\right)$ with pdf and c.d.f. function

$$\begin{aligned}
\phi\left(z\right) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right) \\
\Phi\left(z\right) &= \int_{-\infty}^{z}\phi\left(u\right)du,
\end{aligned}$$

Then, the conditional density function of conditional random variable $z|\, z>c$ (say h) is

$$h\left(z|\, z>c\right) = \frac{\phi\left(z\right)}{1-\Phi\left(c\right)}$$

For detail, see footnote[3].

[3] Assume that f and F are pdf and cdf of r.v. X and x is realization of r.v. Definition of pdf is

$$f\left(x\right) = \frac{d}{dx}F\left(x\right).$$

Then, conditional expectation $E_{z|z>c}\left[z|\, z>c\right]$ is calculated by

$$
\begin{aligned}
& E_{z|z>c}\left[z|\, z>c\right] \\
= & \int_c^{\infty} z\cdot h\left(z|\, z<c\right) dz \quad \text{(definition of conditional expectation)} \\
= & \int_c^{\infty} z\frac{\phi\left(z\right)}{1-\Phi\left(c\right)} dz \\
= & \frac{1}{1-\Phi\left(c\right)}\int_c^{\infty} z\phi\left(z\right) dz \\
= & \frac{1}{\Phi\left(c\right)}\int_c^{\infty} \underbrace{z\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)}_{\text{using trick}} dz \\
= & \frac{1}{1-\Phi\left(c\right)}\int_c^{\infty} \frac{d}{dz}\left\{-\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\right\} dz \quad \left(\text{since } \frac{d}{dz}\left\{-\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\right\} = z\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\right) \\
= & \frac{1}{1-\Phi\left(c\right)}\int_{\infty}^{c} \frac{d}{dz}\left\{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\right\} dz \quad \left(\text{changing integrating region } \int_a^b -f = \int_b^a f\right) \\
= & \frac{1}{1-\Phi\left(c\right)}\frac{d}{dc}\int_{\infty}^{c} \underbrace{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)}_{=\phi(z)} dz \quad \left(\text{exchanging } \int_{\infty}^{c} \text{ and } \frac{d}{dz} \text{ with changing variables}\right) \\
= & \frac{1}{1-\Phi\left(c\right)}\frac{d}{dc}\int_{\infty}^{c} \phi\left(u\right) du \\
= & \frac{\phi\left(c\right)}{1-\Phi\left(c\right)} \quad \text{(by fundamental theorem of calculus see footnote)}
\end{aligned}
$$

i.e. we have

$$
E_{z|z>c}\left[z|\, z>c\right] = \frac{\phi\left(c\right)}{1-\Phi\left(c\right)} \tag{4}
$$

Now, we apply the result above. Since we have the assumption $\varepsilon_i|\, x_i \sim N\left(0_{1\times 1}, \sigma^2\right)$, we can calculate conditional

---

Assuming that c is a constant. Then, the densigy function of $f\left(x|\, x>c\right)$ is

$$
\begin{aligned}
f\left(x|\, x>c\right) &= \frac{d}{dx}F\left(x|\, x>c\right) = \frac{d}{dx}\Pr\left(X<x|\, x>c\right) \\
&= \frac{d}{dx}\frac{\Pr\left(X<x \text{ and } x>c\right)}{\Pr\left(x>c\right)} \\
&= \frac{\frac{d}{dx}\int_c^x f\left(u\right) du}{1-F\left(c\right)} \quad \left(\text{since } \Pr\left(x>c\right) = 1-F\left(c\right)\right) \\
&= \frac{f\left(x\right)}{\Pr\left(x>c\right)} \quad \text{(by fundamentl theorem of calculus)}
\end{aligned}
$$

Fundamentl theorem of calculus:
If we define the function G and g on the interval $[a,b]$ by

$$
G\left(x\right) = \int_a^x g\left(u\right) du,
$$

then we have

$$
\frac{d}{dx}G\left(x\right) = g\left(x\right)
$$

expectation

$$\begin{aligned}
& E_{\varepsilon_i|x_i'\beta>-\varepsilon}\left[\left.\underbrace{\varepsilon_i}_{\text{distributed as } N(0_{1\times1},\sigma^2)}\right|\varepsilon>-x_i'\beta\right] \\
= \;& \sigma\cdot E_{\varepsilon|\varepsilon>-x_i'\beta}\left[\left.\frac{\varepsilon_i}{\sigma}\right|\varepsilon>-x_i'\beta\right] \qquad (\sigma\text{'s are created}) \\
= \;& \sigma\cdot E_{\varepsilon|\frac{\varepsilon}{\sigma}>-\frac{x_i'\beta}{\sigma}}\left[\left.\underbrace{\frac{\varepsilon_i}{\sigma}}_{\text{distributed as standard normal}}\right|\frac{\varepsilon}{\sigma}>-\frac{x_i'\beta}{\sigma}\right] \\
= \;& \sigma\cdot\left[\frac{\phi\left(-\frac{x_i'\beta}{\sigma}\right)}{1-\Phi\left(-\frac{x_i'\beta}{\sigma}\right)}\right] \qquad (\text{applying the result (4)}) \\
= \;& \sigma\cdot\left[\frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{1-\left[1-\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]}\right] \qquad (\text{since } \phi(-z)=\phi(z) \text{ and } \Phi(-z)=1-\Phi(z)\,) \\
= \;& \sigma\cdot\frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}.
\end{aligned}$$

So, we go the conditional expectations

$$E_{y_i|x_i,y_i>0}\left[y_i|\,x_i,y_i>0\right]=x_i'\beta+\sigma\cdot\frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}. \tag{5}$$

Also, from calculation before

$$\begin{aligned}
E_{y_i|x_i}\left[y_i|\,x_i\right] &= \Pr\left(y^*>0|\,x_i\right)\cdot E_{y_i|x_i,y_i>0}\left[y_i|\,x_i,y_i>0\right] \qquad (\#) \\
&= \Phi\left(\frac{x_i'\beta}{\sigma}\right)\cdot\left[x_i'\beta+\sigma\cdot\frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}\right] \\
&= \Phi\left(\frac{x_i'\beta}{\sigma}\right)x_i'\beta+\phi\left(\frac{x_i'\beta}{\sigma}\right)
\end{aligned}$$

In summary, we got two conditional expectations

$$E_{y_i|x_i,y_i>0}\left[y_i|\,x_i,y_i>0\right]=x_i'\beta+\sigma\cdot\frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)} \tag{6}$$

and

$$E_{y_i|x_i}\left[y_i|\,x_i\right]=\Phi\left(\frac{x_i'\beta}{\sigma}\right)x_i'\beta+\phi\left(\frac{x_i'\beta}{\sigma}\right) \tag{7}$$

Notice that the conditional expectation (7) is the conditional expectation that consist of all y_i (especially, includes $y_i=0$ data). On the other hand, the conditional expectation (6) consists of only $y_i>0$ data (and does not includes). Therefore, we have the relation

$$E_{x_i,y_i>0}\left[y_i|\,x_i,y_i>0\right]>E_{y_i|x_i}\left[y_i|\,x_i\right].$$

Actually, this relation if obvious from equation (#), because $\Pr\left(y^*>0|\,x_i\right)$ is between 0 and 1.

2.4 Heckman Two-Step Estimation

The basic estimation method of Tobit model is maximum likelihood. However, we have a very simple alternative estimation method, Heckman two-step estimation.

Based on conditional expectation (5), we can implement Hackman two-step estimation for obtaining consistent estimator of β and σ.

Define the hazard (inverse Mill) function

$$\begin{aligned}\lambda(z) &= \frac{\phi(z)}{\Phi(z)} \\ \lambda\left(\frac{x_i'\beta}{\sigma}\right) &= \frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}.\end{aligned}$$

Then, (6) can be written as

$$E\left[y_i|\, x_i, y_i > 0\right] = x_i'\beta + \sigma \cdot \lambda\left(\frac{x_i'\beta}{\sigma}\right).$$

Based on above equation, Heckman constructed the model equation for data that satisfy $y_i > 0$

$$y_i = x_i'\beta + \sigma \cdot \lambda\left(\frac{x_i'\beta}{\sigma}\right) + u_i$$

where the error term u_i has zero conditional expectation

$$E\left[u_i|\, x_i, y_i > 0\right] = 0.$$

We implement estimation with two steps.

Step 1: Using all data and implement probit (or logit) estimation by using dummy variable (rewriting the definition of dummy variable d_i)

$$d_i = \begin{cases} 1 & : \text{ if } y_i > 0 \\ 0 & : \text{ if } y_i = 0 \end{cases},$$

and construct the probit (or logit) model

$$\begin{aligned}d_i &= x_i'\beta + \varepsilon_i \\ \underbrace{d_i}_{0 \text{ or } 1} &= x_i'\beta + \underbrace{\varepsilon_i}_{\text{distributed as } N(0,\sigma^2)}.\end{aligned}$$

Then, we can estimate (you know, in binary choice model, we can estimate β up to scale)

$$\widehat{\left(\frac{\beta}{\sigma}\right)},$$

by probit (or logit) estimation.

Step 2: Calculate the estimate of hazard function by using $\widehat{\left(\frac{\beta}{\sigma}\right)}$ in step 1

$$\lambda\widehat{\left(\frac{x_i'\beta}{\sigma}\right)} = \lambda\left(x_i'\widehat{\left(\frac{\beta}{\sigma}\right)}\right) = \frac{\phi\left(x_i'\widehat{\left(\frac{\beta}{\sigma}\right)}\right)}{\Phi\left(x_i'\widehat{\left(\frac{\beta}{\sigma}\right)}\right)}.$$

Then, implement OLS for the model by only using $y_i > 0$ data

$$y_i = x_i'\beta + \sigma \cdot \lambda\widehat{\left(\frac{x_i'\beta}{\sigma}\right)} + u_i,$$

and obtain estimator $\hat{\beta}$ and $\hat{\sigma}$.

The advantage of Heckman two-step estimation is its simplicity. We only need the program of probit (or logit) estimation and OLS estimation.

3 Comp 2004S Part II (Kyriazidou) Question 2: Type I Standard Tobit Model

Suppose that we have an i.i.d. sample of N observations on (y_i, x_i) from censored regression model

$$y_i = \max\{y_i^*, 0\}$$

where the latent variable y_i^* is given by a linear regression model

$$y_i^* = x_i'\beta_0 + \sigma_0\varepsilon_i$$

where $\sigma_0 > 0$ is known and ε_i is distributed according to a known continuous distribution function with density function f independent of x_i.

(1) Define the maximum likelihood estimator of β_0.

Answer:

Notice that this question asks that we need to "define" the MLE of β_0, and we just need to characterize the ML estimator.

First of all, calculating the conditional probabilities.

$$\begin{aligned}
\Pr(y_i > 0|x_i) &= \Pr(y_i^* > 0|x_i) = \Pr(x_i'\beta_0 + \sigma_0\varepsilon_i > 0|x_i) = \Pr\left(\varepsilon_i > -\frac{x_i'\beta_0}{\sigma_0}\middle| x_i\right) = 1 - F\left(-\frac{x_i'\beta_0}{\sigma_0}\right) \\
\Pr(y_i = 0|x_i) &= 1 - \Pr(y_i^* > 0|x_i) = 1 - \left\{1 - F\left(-\frac{x_i'\beta_0}{\sigma_0}\right)\right\} = F\left(-\frac{x_i'\beta_0}{\sigma_0}\right)
\end{aligned}$$

Notice that we do not have symmetric distribution, so

$$1 - F\left(-\frac{x_i'\beta_0}{\sigma_0}\right) \neq F\left(\frac{x_i'\beta_0}{\sigma_0}\right).$$

Define that

$$d_i = \begin{cases} 1 & \text{if} \quad y_i > 0 \quad \text{positive values are observed} \\ 0 & \text{if} \quad y_i = 0 \quad \text{otherwise} \end{cases}.$$

Then, the conditional density function is

$$g(y_i; \beta|x_i) = \{f(y_i|x_i)\}^{d_i}\left\{F\left(-\frac{x_i'\beta_0}{\sigma_0}\right)\right\}^{1-d_i}$$

The n-sample likelihood function is

$$L_n(\beta) = \prod_{i=1}^{n} g(y_i; \beta|x_i) \qquad (\text{since } (y_i, x_i) \text{ are i.i.d.})$$

The n-sample log-likelihood function is

$$\begin{aligned}
l_n(\beta) &= \ln L = \ln\prod_{i=1}^{n} g(y_i, x_i; \beta) = \sum_{i=1}^{n}\ln g(y_i; \beta|x_i) \\
&= \sum_{i=1}^{n}\ln\left\{[f(y_i|x_i)]^{d_i}\left[F\left(-\frac{x_i'\beta_0}{\sigma_0}\right)\right]^{1-d_i}\right\} \\
&= \sum_{i=1}^{n}\left\{d_i\ln f(y_i|x_i) + (1-d_i)\ln F\left(-\frac{x_i'\beta_0}{\sigma_0}\right)\right\}.
\end{aligned}$$

The maximum likelihood estimator is defined as

$$\hat{\beta}_{ML} = \arg\max_{\beta}\{l_n(\beta)\}.$$

(2) Derive the conditional mean of y_i given x_i, i.e. $E[y_i|x_i]$

Answer:

Actually, to solve this question, we have to specify pdf f. So, we assume f is pdf of normal distribution. Then, derivation is the same as subsection 2.3. So, we skip the derivation (see section 2.3) and conditional expectation (mean) is

$$E_{y_i|x_i}\left[y_i|\, x_i\right] = \Phi\left(\frac{x_i'\beta}{\sigma}\right)\cdot\left[x_i'\beta + \sigma\cdot\lambda\left(\frac{x_i'\beta}{\sigma}\right)\right]$$

where λ is hazard (inverse mill) function

$$\lambda\left(z\right) = \frac{\phi\left(z\right)}{\Phi\left(z\right)}.$$

(3) Derive the marginal effect on x_i on $E\left[y_i|x_i\right]$

Answer:

Then, the marginal effect is

$$\begin{aligned}
\frac{\partial}{\partial x_i}E_{y_i|x_i}\left[y_i|x_i\right] &= \frac{\partial}{\partial x_i}\left[\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\cdot\left\{x_i'\beta_0+\sigma_0\lambda\left(\frac{x_i'\beta_0}{\sigma_0}\right)\right\}\right] && \text{(from (result in (2))}\\
&= \frac{\partial}{\partial x_i}\left[\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\cdot\left\{x_i'\beta_0+\sigma_0\frac{\phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)}{\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)}\right\}\right] && \text{(since } \lambda\left(z\right)=\frac{\phi\left(z\right)}{\Phi\left(z\right)}\text{)}\\
&= \frac{\partial}{\partial x_i}\left[\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)x_i'\beta_0+\sigma_0\cdot\phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\right]\\
&= \underbrace{\phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\frac{\beta_0}{\sigma_0}x_i'\beta_0+\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\beta_0}_{\text{product rule}}+\sigma_0\cdot\underbrace{\phi'\left(\frac{x_i'\beta_0}{\sigma_0}\right)}_{=-\frac{x_i'\beta_0}{\sigma_0}\phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\text{ see below}}\frac{\beta_0}{\sigma_0}\\
&= \underbrace{\phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\frac{\beta_0}{\sigma_0}x_i'\beta_0}_{\text{cancel out}}+\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\beta_0-\underbrace{\sigma_0\frac{x_i'\beta_0}{\sigma_0}\phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\frac{\beta_0}{\sigma_0}}_{\text{cancel out}}\\
&= \Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\beta_0
\end{aligned}$$

where we used the fact

$$\phi'\left(z\right)=\frac{d}{dz}\left\{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\right\}=-z\underbrace{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)}_{=\phi(z)}=-z\phi\left(z\right)$$

(4) Evaluate the claim: "Censoring leads to an attenuation of the marginal effect of X_i relative to its effect in the latent regression."

Answer:

From the result in (3), we have

$$\frac{\partial}{\partial x_i}E_{y_i|x_i}\left[y_i|x_i\right]=\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)\beta_0\le\beta_0 \qquad \text{(since } 0\le\Phi\left(\cdot\right)\le 1\text{)}$$

Thus, we can conclude censoring leads weakening the effect of x_{ik}. This is intuitive from the graph we wrote in subsection 2.1. (see graph)

4 Final 2003: Question 4: Type I (Top-Coded) Tobit Model

In order to be able to assess the performance of firms, data was collected on the expenditure of firms y^* and other variables x_i, $i = 1, \ldots, n$. Unfortunately the data was top-coded, so that if the expenditure for any firm was above y_0 it was recorded as y_0. Suppose that the model is linear, that is

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

where

$$E\left[\varepsilon_i | x_i\right] = 0.$$

Let $y_i = y_i^*$ if $y_i^* < y_0$ and let $y_i = y_0$ if $y_i^* \geq y_0$.

Preparation:

Since we cannot derive closed form solution of conditional expectation and apply Heckman two step estimation without normality assumption of error terms, we assume

$$\varepsilon_i \sim N\left(0, \sigma^2\right).$$

Then, (even though Moshe does not ask you to derive ML estimator, it is useful to discuss), the density and likelihood function are

$$f\left(y_i | x_i\right) = \begin{cases} \text{mass part} & \Pr\left(y_i = y_0 | x_i\right) = \Phi\left(\frac{y_0 - x_i'\beta}{\sigma}\right) & : \text{ if } y_i \leq 0 \\ \text{continuous part} & f^*\left(y_i | x_i\right) = \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) & : \text{ if } y_i > 0 \end{cases}$$

See foot note[4].

Next, define the dummy variable

$$d_i = \begin{cases} 1 & : \text{ if } y_i > 0 \\ 0 & : \text{ if } y_i \leq 0 \end{cases}.$$

Then, by (2) and (3), we can express the conditional pdf of y_i given x_i as

$$\begin{aligned} f\left(y_i | x_i\right) &= \left\{\Pr\left(y_i = y_0 | x_i\right)\right\}^{d_i} \cdot \left\{f^*\left(y_i | x_i\right)\right\}^{1-d_i} \\ &= \left\{\Phi\left(\frac{y_0 - x_i'\beta}{\sigma}\right)\right\}^{d_i} \cdot \left\{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\}^{1-d_i}. \end{aligned}$$

and likelihood is defined as

$$L_n\left(\beta, \sigma^2\right) = \prod_{i=1}^{N} f\left(y_i | x_i\right).$$

Also, log-likelihood is define as

$$\begin{aligned} l_n\left(\beta, \sigma^2\right) &= \ln L_n\left(\beta, \sigma^2\right) \\ &= \sum_{i=1}^{N} \ln f\left(y_i | x_i\right) \\ &= \sum_{i=1}^{N} \ln\left[\left\{\Phi\left(\frac{y_0 - x_i'\beta}{\sigma}\right)\right\}^{d_i} \cdot \left\{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\}^{1-d_i}\right] \\ &= \sum_{i=1}^{N}\left[d_i \ln\left\{\Phi\left(\frac{y_0 - x_i'\beta}{\sigma}\right)\right\} + (1 - d_i)\left\{-\ln\sigma + \ln\left\{\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\}\right\}\right]. \end{aligned}$$

The ML estimator is defined as

$$\left(\beta, \sigma^2\right) = \arg\max_{\beta, \sigma^2}\left\{l_n\left(\beta, \sigma^2\right)\right\}.$$

---

[4] Mass part is derived by

$$\begin{aligned} \Pr\left(y_i = y_0 | x_i\right) &= \Pr\left(y_i^* \geq y_0 | x_i\right) = \Pr\left(x_i'\beta + \varepsilon_i \geq y_0 | x_i\right) \\ &= \Pr\left(\varepsilon_i \geq y_0 - x_i'\beta | x_i\right) = \Pr\left(\left.\underbrace{\frac{\varepsilon_i}{\sigma_0}}_{\text{distributed standard normally}} \geq \frac{y_0 - x_i'\beta}{\sigma_0}\right| x_i\right) \\ &= \Phi\left(\frac{y_0 - x_i'\beta}{\sigma_0}\right). \end{aligned}$$

Graph:

(1) Compute $E\left[y_i | x_i, y_i < y_0\right]$.
Answer:

$$
\begin{aligned}
& E_{y_i|x_i,y_i<y_0}\left[y_i | x_i, y_i < y_0\right] \\
= \; & E_{\varepsilon_i|x_i,y_i<y_0}\left[y_i | x_i, y_i < y_0\right] \\
= \; & E_{\varepsilon_i|x,y^*<y_0}\left[y^* | x, y^* < y_0\right] \qquad \text{(since } y_i = y_i \text{ if } y_i < y_0\text{)} \\
= \; & E_{\varepsilon_i|x_i,x_i'\beta+\varepsilon_i<y_0}\left[x_i'\beta + \varepsilon_i | x_i, x_i'\beta + \varepsilon < y_0\right] \\
= \; & E_{\varepsilon_i|,x_i'\beta+\varepsilon<y_0}\left[x_i'\beta + \varepsilon_i | x_i'\beta + \varepsilon < y_0\right] \qquad \text{(deleting redundant notation)} \\
= \; & E_{\varepsilon_i|\varepsilon_i<y_0-x_i'\beta}\left[x_i'\beta + \varepsilon_i | \varepsilon_i < y_0 - x_i'\beta\right] \\
= \; & x_i'\beta + E_{\varepsilon|\varepsilon<y_0-x_i'\beta}\left[\varepsilon_i | \varepsilon_i < y_0 - x_i'\beta\right]
\end{aligned}
$$

Assume that if $z \sim N\left(0_{1\times1}, 1\right)$ with pdf and c.d.f. function

$$
\begin{aligned}
\phi(z) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \\
\Phi(z) &= \int_{-\infty}^{z} \phi(u)\, du,
\end{aligned}
$$

Then, the conditional density function of conditional random variable $z | z < c$ (say h) is

$$
h\left(z | z < c\right) = \frac{\phi(z)}{\Phi(c)}
$$

See foot note for derivation above[5].

[5] The derivation is

$$
h\left(z | z < c\right) = \frac{d}{dz} \Pr\left(z | z < c\right) = \frac{d}{dz} \frac{\Pr\left(z \cap z < c\right)}{\Pr\left(z < c\right)} = \frac{\frac{d}{dz}\int_{-\infty}^{z \text{ such that } z<c} f(u)\, du}{\Phi(c)} = \frac{f(z)}{\Phi(c)} \quad \text{(if } z < c\text{)}.
$$

Then, conditional expectation $E_{z|z<c}\left[z|\,z<c\right]$ is calculated as

$$
\begin{aligned}
& E_{z|z<c}\left[z|\,z<c\right] \\
= & \int_{-\infty}^{c} z\cdot h\left(z|\,z<c\right)dz \quad \text{(definition of conditional expectation)} \\
= & \int_{-\infty}^{c} z\frac{\phi\left(z\right)}{\Phi\left(c\right)}dz \quad \text{(since } h\left(z|\,z<c\right)=\frac{\phi\left(z\right)}{\Phi\left(c\right)}\text{)} \\
= & \frac{1}{\Phi\left(c\right)}\int_{-\infty}^{c} z\phi\left(z\right)dz=\frac{1}{\Phi\left(c\right)}\int_{-\infty}^{c}\underbrace{z\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^{2}\right)}_{\text{using trick}}dz \\
= & \frac{1}{\Phi\left(c\right)}\int_{\infty}^{c}\frac{d}{dz}\left\{-\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^{2}\right)\right\}dz \quad \text{(since } \frac{d}{dz}\left\{-\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^{2}\right)\right\}=z\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^{2}\right)\text{)} \\
= & \frac{-1}{\Phi\left(c\right)}\int_{\infty}^{c}\frac{d}{dz}\left\{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^{2}\right)\right\}dz \\
= & \frac{-1}{\Phi\left(c\right)}\frac{d}{dc}\int_{\infty}^{c}\left\{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^{2}\right)\right\}dz \quad \text{(exchanging } \int_{\infty}^{c} \text{ and } \frac{d}{dz} \text{ with changing variables)} \\
= & -\frac{\phi\left(c\right)}{\Phi\left(c\right)} \quad \text{(by fundamental theorem of calculus)}
\end{aligned}
$$

Since we have the assumption $\varepsilon_i|\,x_i\sim N\left(0_{1\times 1},\sigma^2\right)$, we can calculate conditional probability

$$
\begin{aligned}
E_{\varepsilon|\varepsilon<y_0-x_i'\beta}\left[\varepsilon_i|\,\varepsilon_i<y_0-x_i'\beta\right] &= \sigma\cdot E_{\varepsilon_i|\varepsilon_i<y_0-x_i'\beta+y_0}\left[\left.\frac{\varepsilon_i}{\sigma}\right|\varepsilon_i<y_0-x_i'\beta\right] \qquad (\sigma\text{'s are created}) \\
&= \sigma\cdot E_{\frac{\varepsilon_i}{\sigma}|\frac{\varepsilon_i}{\sigma}<\frac{y_0-x_i'\beta}{\sigma}}\left[\underbrace{\frac{\varepsilon_i}{\sigma}}_{\text{distributed standard}}\left|\frac{\varepsilon}{\sigma}<\frac{y_0-x_i'\beta}{\sigma}\right.\right] \\
&= \sigma\cdot\left[-\frac{\phi\left(\frac{y_0-x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{y_0-x_i'\beta}{\sigma}\right)}\right] \qquad \text{(by using result above } E_{z|z<c}\left[z|\,z<c\right]=-\frac{\phi\left(c\right)}{\Phi\left(c\right)}\text{)} \\
&= -\sigma\cdot\left[\frac{\phi\left(\frac{y_0-x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{y_0-x_i'\beta}{\sigma}\right)}\right]
\end{aligned}
$$

So, we go the conditional expectations

$$
E_{y_i|x_i,y_i>0}\left[y_i|\,x_i,y_i>0\right]=x_i'\beta-\sigma\cdot\left[\frac{\phi\left(\frac{y_0-x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{y_0-x_i'\beta}{\sigma}\right)}\right]
$$

or equivalently,

$$
E_{y_i|x_i,y_i>0}\left[y_i|\,x_i,y_i>0\right]=x_i'\beta-\sigma\cdot\lambda\left(\frac{y_0-x_i'\beta}{\sigma}\right),
$$

where λ is hazard (inverse Mill) function defined by

$$
\lambda\left(z\right)=\frac{\phi\left(z\right)}{\Phi\left(z\right)}.
$$

(2) Propose a two-step method for estimating β.

Answer:

Step 1: Using all data and implement probit (or logit) estimation by using dummy variable

$$
d_i=\begin{cases} 1 & : \text{ if } y_i=y_0 \\ 0 & : \text{ if } y_i<y_0 \end{cases},
$$

and construct the probit (or logit) model

$$\begin{aligned} d_i &= x_i'\beta + \varepsilon_i \\ \underbrace{d_i}_{0 \text{ or } 1} &= x_i'\beta + \underbrace{\varepsilon_i}_{\text{distributed as } N(0,\sigma^2)}. \end{aligned}$$

Then, we can estimate (in binary choice model, we can estimate β up to scale)

$$\widehat{\left(\frac{\beta}{\sigma}\right)},$$

by probit (or logit) estimation.

Step 2: Calculate the estimate of hazard function by using $\widehat{\left(\frac{\beta}{\sigma}\right)}$ in step 1

$$\lambda\left(\widehat{\frac{y_0 - x_i'\beta}{\sigma}}\right) = \lambda\left(\frac{y_0}{\hat{\sigma}} - x_i'\widehat{\left(\frac{\beta}{\sigma}\right)}\right) = \frac{\phi\left(\frac{y_0}{\hat{\sigma}} - x_i'\widehat{\left(\frac{\beta}{\sigma}\right)}\right)}{\Phi\left(\frac{y_0}{\hat{\sigma}} - x_i'\widehat{\left(\frac{\beta}{\sigma}\right)}\right)}.$$

Notice that we need some prior assumption of $\hat{\sigma}$ to calculate $\lambda\widehat{\left(\frac{x_i'\beta}{\sigma}\right)}$ (alternative method is repeating step 1 and step 2 until parameter converses)

Then, implement OLS for the model by only using $y_i > 0$ data

$$y_i = x_i'\beta - \sigma \cdot \lambda\left(\widehat{\frac{y_0 - x_i'\beta}{\sigma}}\right) + u_i,$$

and obtain estimator $\hat{\beta}$ and $\hat{\sigma}$..

(3) Suppose now that each firm had a different, but known, top-coding value, given by $y_{i,0}$, for $i = 1, \ldots, n$. Will the method proposed in (2) apply here as well? Explain

Answer:

If we know different top-coded (censoring) values $y_{i,0}$, we still can construct density and likelihood function with density functions

$$f\left(y_i | x_i\right) = \left\{\Phi\left(\frac{\overbrace{y_{i,0}}^{\text{varying}} - x_i'\beta}{\sigma}\right)\right\}^{d_i} \cdot \left\{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right\}^{1-d_i}.$$

(4) Show that the method proposed in (2) for estimating β can be viewed as GMM method.

Answer:

From (2), in the step 2 of Heckman two-step estimation, we regress (by OLS) the equation with $y_i < y_0$ data

$$\begin{aligned} y_i &= x_i'\beta - \sigma \cdot \lambda\left(\frac{y_0 - x_i'\beta}{\sigma}\right) + u_i, \\ u_i &= y_i - x_i'\beta + \sigma \cdot \lambda\left(\frac{y_0 - x_i'\beta}{\sigma}\right) \end{aligned}$$

where

$$E\left[u_i | x_i, y_i < y_0\right] = 0.$$

Denote $w_i = [y_i, x_i]$

Define the function

$$\underbrace{\varphi_1\left(w_i, \beta, \sigma\right)}_{(K+1)\times 1} = \left\{y_i - x_i'\beta + \sigma \cdot \lambda\left(\frac{y_0 - x_i'\beta}{\sigma}\right)\right\}\begin{bmatrix} x_i \\ \lambda\left(\frac{y_0 - x_i'\beta}{\sigma}\right) \end{bmatrix}.$$

Then, we have the population moment condition

$$\begin{aligned} E_{u_i,x_i}\left[\varphi_1\left(w_i,\beta,\sigma\right)\right] &= E_{u_i,x_i}\left[\left\{y_i - x_i'\beta + \sigma\cdot\lambda\left(\frac{y_0 - x_i'\beta}{\sigma}\right)\right\}\left[\begin{array}{c} x_i \\ \lambda\left(\frac{y_0-x_i'\beta}{\sigma}\right)\end{array}\right]\middle| x_i, y_i < y_0\right] \\ &= E_{u_i,x_i}\left[u_i\left[\begin{array}{c} x_i \\ \lambda\left(\frac{y_0-x_i'\beta}{\sigma}\right)\end{array}\right]\middle| x_i, y_i < y_0\right] = E_{x_i}\left[\underbrace{E_{u_i|x_i}\left[u_i\right]}_{=0}\cdot\left[\begin{array}{c} x_i \\ \lambda\left(\frac{y_0-x_i'\beta}{\sigma}\right)\end{array}\right]\middle| x_i, y_i < y_0\right] = 0_{(K+1)\times 1} \end{aligned}$$

Thus, we have the sample analogue of moment

$$m_{n,1}\left(\beta,\sigma\right) = \frac{1}{N}\sum_{i=1}^{N}\left[\left\{y_i - x_i'\beta + \sigma\cdot\lambda\left(\frac{y_0 - x_i'\beta}{\sigma}\right)\right\}\left[\begin{array}{c} x_i \\ \lambda\left(\frac{y_0-x_i'\beta}{\sigma}\right)\end{array}\right]\right] = 0_{(K+1)\times 1},$$

MM estimator is defined by (numerically) solving above system equations. Notice that we only use data such that $y_i < y_0$ for this estimation.

(5) Propose a GMM method (but not an MLE) that will provide a better estimator than one proposed in (2).

Answer:

Define the function (for example) with

$$\underbrace{\varphi_2\left(w_i,\beta,\sigma\right)}_{(2(K+1))\times 1} = \left\{y_i - x_i'\beta + \sigma\cdot\lambda\left(\frac{y_0 - x_i'\beta}{\sigma}\right)\right\}\left[\begin{array}{c} x_i \\ \lambda\left(\frac{y_0-x_i'\beta}{\sigma}\right) \\ x_i.\hat{}2 \\ \left[\lambda\left(\frac{y_0-x_i'\beta}{\sigma}\right)\right]^2\end{array}\right].$$

Then, we have the moment condition

$$E_{u_i,x_i}\left[\varphi_2\left(w_i,\beta,\sigma\right)\right] = 0_{(2(K+1))\times 1.}$$

Define sample analogue,

$$m_{n,2}\left(\beta,\sigma\right) = \frac{1}{N}\sum_{i=1}^{N}\varphi_2\left(w_i,\beta,\sigma\right).$$

Then, implement optimal GMM estimation and we can obtain optimal GMM estimator. Note that compared to moment condition in (4), moment condition in (5) has more moments and can capture more error structures. Thus, estimator in (5) is more efficient. Again, notice that we only use data such that $y_i < y_0$ for this estimation.