2.2 Limited Dependent Variable Model (制限従属変数モデル)

Truncated Regression Model: Consider the following model:

$$y_i = X_i\beta + u_i, \qquad u_i \sim N(0, \sigma^2) \text{ when } y_i > a, \text{ where } a \text{ is a constant},$$

for $i = 1, 2, \cdots, n$.

Consider the case of $y_i > a$ (i.e., in the case of $y_i \leq a$, y_i is not observed).

$$\mathrm{E}(u_i|X_i\beta + u_i > a) = \int_{a-X_i\beta}^{\infty} u_i \frac{f(u_i)}{1 - F(a - X_i\beta)} \mathrm{d}u_i.$$

Suppose that $u_i \sim N(0, \sigma^2)$, i.e., $\dfrac{u_i}{\sigma} \sim N(0, 1)$.

Using the following standard normal density and distribution functions:

$$\phi(x) = (2\pi)^{-1/2} \exp(-\frac{1}{2}x^2),$$

$$\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-\frac{1}{2}z^2)\mathrm{d}z = \int_{-\infty}^{x} \phi(z)\mathrm{d}z,$$

$f(x)$ and $F(x)$ are given by:

$$f(x) = (2\pi\sigma^2)^{-1/2}\exp(-\frac{1}{2\sigma^2}x^2) = \frac{1}{\sigma}\phi(\frac{x}{\sigma}),$$
$$F(x) = \int_{-\infty}^{x}(2\pi\sigma^2)^{-1/2}\exp(-\frac{1}{2\sigma^2}z^2)\mathrm{d}z = \Phi(\frac{x}{\sigma}).$$

[Review — Mean of Truncated Normal Random Variable:]

Let X be a normal random variable with mean μ and variance σ^2.

Consider $\mathrm{E}(X|X > a)$, where a is known.

The truncated distribution of X given $X > a$ is:

$$f(x|x > a) = \frac{(2\pi\sigma^2)^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big)}{\int_a^{\infty}(2\pi\sigma^2)^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big)\mathrm{d}x} = \frac{\frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})}{1-\Phi(\frac{a-\mu}{\sigma})}.$$

$$E(X|X>a) = \int_a^\infty x f(x|x>a)\mathrm{d}x = \frac{\int_a^\infty x(2\pi\sigma^2)^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big)\mathrm{d}x}{\int_a^\infty (2\pi\sigma^2)^{-1/2}\exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big)\mathrm{d}x}$$

$$= \frac{\sigma\phi(\frac{a-\mu}{\sigma}) + \mu\Big(1-\Phi(\frac{a-\mu}{\sigma})\Big)}{1-\Phi(\frac{a-\mu}{\sigma})} = \frac{\sigma\phi(\frac{a-\mu}{\sigma})}{1-\Phi(\frac{a-\mu}{\sigma})} + \mu,$$

which are shown below. The denominator is:

$$\int_a^\infty (2\pi\sigma^2)^{-1/2}\exp(-\frac{1}{2\sigma^2}(x-\mu)^2)\mathrm{d}x = \int_{\frac{a-\mu}{\sigma}}^\infty (2\pi)^{-1/2}\exp(-\frac{1}{2}z^2)\mathrm{d}z$$

$$= 1 - \int_{-\infty}^{\frac{a-\mu}{\sigma}} (2\pi)^{-1/2}\exp(-\frac{1}{2}z^2)\mathrm{d}z$$

$$= 1 - \Phi(\frac{a-\mu}{\sigma}),$$

where x is transformed into $z = \frac{x-\mu}{\sigma}$. $\qquad x > a \implies z = \frac{x-\mu}{\sigma} > \frac{a-\mu}{\sigma}$.

The numerator is:

$$
\begin{aligned}
&\int_a^\infty x(2\pi\sigma^2)^{-1/2}\exp(-\frac{1}{2\sigma^2}(x-\mu)^2)\mathrm{d}x \\
&\quad = \int_{\frac{a-\mu}{\sigma}}^\infty (\sigma z+\mu)(2\pi)^{-1/2}\exp(-\frac{1}{2}z^2)\mathrm{d}z \\
&\quad = \sigma\int_{\frac{a-\mu}{\sigma}}^\infty z(2\pi)^{-1/2}\exp(-\frac{1}{2}z^2)\mathrm{d}z + \mu\int_{\frac{a-\mu}{\sigma}}^\infty (2\pi)^{-1/2}\exp(-\frac{1}{2\sigma^2}z^2)\mathrm{d}z \\
&\quad = \sigma\int_{\frac{1}{2}(\frac{a-\mu}{\sigma})^2}^\infty (2\pi)^{-1/2}\exp(-t)\mathrm{d}t + \mu\Big(1-\Phi(\frac{a-\mu}{\sigma})\Big) \\
&\quad = \sigma\phi(\frac{a-\mu}{\sigma}) + \mu\Big(1-\Phi(\frac{a-\mu}{\sigma})\Big),
\end{aligned}
$$

where z is transformed into $t=\frac{1}{2}z^2$. $\quad z>\frac{a-\mu}{\sigma} \implies t=\frac{1}{2}z^2>\frac{1}{2}(\frac{a-\mu}{\sigma})^2$.

[End of Review]

Therefore, the conditional expectation of u_i given $X_i\beta + u_i > a$ is:

$$\mathrm{E}(u_i|X_i\beta + u_i > a) = \int_{a-X_i\beta}^{\infty} u_i \frac{f(u_i)}{1 - F(a - X_i\beta)} \mathrm{d}u_i = \int_{a-X_i\beta}^{\infty} \frac{u_i}{\sigma} \frac{\phi(\dfrac{u_i}{\sigma})}{1 - \Phi(\dfrac{a - X_i\beta}{\sigma})} \mathrm{d}u_i$$

$$= \frac{\sigma\phi(\dfrac{a - X_i\beta}{\sigma})}{1 - \Phi(\dfrac{a - X_i\beta}{\sigma})}.$$

Accordingly, the conditional expectation of y_i given $y_i > a$ is given by:

$$\mathrm{E}(y_i|y_i > a) = \mathrm{E}(y_i|X_i\beta + u_i > a) = \mathrm{E}(X_i\beta + u_i|X_i\beta + u_i > a)$$

$$= X_i\beta + \mathrm{E}(u_i|X_i\beta + u_i > a) = X_i\beta + \frac{\sigma\phi(\dfrac{a - X_i\beta}{\sigma})}{1 - \Phi(\dfrac{a - X_i\beta}{\sigma})},$$

for $i = 1, 2, \cdots, n$.

Estimation:

MLE:

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{f(y_i - X_i\beta)}{1 - F(a - X_i\beta)} = \prod_{i=1}^{n} \frac{1}{\sigma} \frac{\phi(\dfrac{y_i - X_i\beta}{\sigma})}{1 - \Phi(\dfrac{a - X_i\beta}{\sigma})}$$

is maximized with respect to β and σ^2.

Some Examples:

1. Buying a Car:

   $y_i = x_i\beta + u_i$, where y_i denotes expenditure for a car, and x_i includes income, price of the car, etc.

   Data on people who bought a car are observed.

   People who did not buy a car are ignored.

2. Working-hours of Wife:

   y_i represents working-hours of wife, and x_i includes the number of children, age, education, income of husband, etc.

3. Stochastic Frontier Model:

   $y_i = f(K_i, L_i) + u_i$, where y_i denotes production, K_i is stock, and L_i is amount of labor.

   We always have $y_i \leq f(K_i, L_i)$, i.e., $u_i \leq 0$.

   $f(K_i, L_i)$ is a maximum value when we input K_i and L_i.

Censored Regression Model or Tobit Model:

$$y_i = \begin{cases} X_i\beta + u_i, & \text{if } y_i > a, \\ a, & \text{otherwise.} \end{cases}$$

The probability which y_i takes a is given by:

$$P(y_i = a) = P(y_i \le a) = F(a) \equiv \int_{-\infty}^{a} f(x)\mathrm{d}x,$$

where $f(\cdot)$ and $F(\cdot)$ denote the density function and cumulative distribution function of y_i, respectively.

Therefore, the likelihood function is:

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} F(a)^{I(y_i=a)} \times f(y_i)^{1-I(y_i=a)},$$

where $I(y_i = a)$ denotes the indicator function which takes one when $y_i = a$ or zero otherwise.

When $u_i \sim N(0, \sigma^2)$, the likelihood function is:

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \Big(\int_{-\infty}^{a} (2\pi\sigma^2)^{-1/2} \exp(-\frac{1}{2\sigma^2}(y_i - X_i\beta)^2) \mathrm{d}y_i \Big)^{I(y_i=a)} \times \Big((2\pi\sigma^2)^{-1/2} \exp(-\frac{1}{2\sigma^2}(y_i - X_i\beta)^2) \Big)^{1-I(y_i=a)},$$

which is maximized with respect to β and σ^2.

2.3 Count Data Model (計数データモデル)

Poisson distribution:

$$\mathrm{P}(X = x) = f(x) = \frac{e^{-\lambda}\lambda^x}{x!},$$

for $x = 0, 1, 2, \cdots$.

In the case of Poisson random variable X, the expectation of X is:

$$\mathrm{E}(X) = \sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} \lambda\frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda\sum_{x'=0}^{\infty} \frac{e^{-\lambda}\lambda^{x'}}{x'!} = \lambda.$$

Remember that $\sum_x f(x) = 1$, i.e., $\sum_{x=0}^{\infty} e^{-\lambda}\lambda^x/x! = 1$.

Therefore, the probability function of the count data y_i is taken as the Poisson distribution with parameter λ_i.

In the case where the explained variable y_i takes 0, 1, 2, $\cdots$ (discrete numbers), assuming that the distribution of y_i is Poisson, the logarithm of λ_i is specified as a linear function, i.e.,

$$\mathrm{E}(y_i) = \lambda_i = \exp(X_i\beta).$$

Note that λ_i should be positive.

Therefore, it is better to avoid the specification: $\lambda = X_i\beta$.

The joint distribution of y_1, y_2, $\cdots$, y_n is:

$$f(y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} = L(\beta),$$

where $\lambda_i = \exp(X_i\beta)$.

The log-likelihood function is:

$$\begin{aligned}\log L(\beta) &= -\sum_{i=1}^{n} \lambda_i + \sum_{i=1}^{n} y_i \log \lambda_i - \sum_{i=1}^{n} \log y_i! \\ &= -\sum_{i=1}^{n} \exp(X_i\beta) + \sum_{i=1}^{n} y_i X_i\beta - \sum_{i=1}^{n} \log y_i!.\end{aligned}$$

The first-order condition is:

$$\frac{\partial \log L(\beta)}{\partial \beta} = -\sum_{i=1}^{n} X_i' \exp(X_i\beta) + \sum_{i=1}^{n} X_i' y_i = 0.$$

$\Longrightarrow$ Nonlinear optimization procedure

[Review] Nonlinear Optimization Procedures:

Note that the Newton-Raphson method (one of the nonlinear optimization procedures) is:

$$\beta^{(j+1)} = \beta^{(j)} - \left(\frac{\partial^2 \log L(\beta^{(j)})}{\partial \beta \partial \beta'}\right)^{-1} \frac{\partial \log L(\beta^{(j)})}{\partial \beta},$$

which comes from the first-order Taylor series expansion around $\beta = \beta^*$:

$$0 = \frac{\partial \log L(\beta)}{\partial \beta} \approx \frac{\partial \log L(\beta^*)}{\partial \beta} + \frac{\partial^2 \log L(\beta^*)}{\partial \beta \partial \beta'}(\beta - \beta^*),$$

and β and β^* are replaced by $\beta^{(j+1)}$ and $\beta^{(j)}$, respectively.

An alternative nonlinear optimization procedure is known as the method of scoring, which is shown as:

$$\beta^{(j+1)} = \beta^{(j)} - \left(\mathrm{E}\Big(\frac{\partial^2 \log L(\beta^{(j)})}{\partial\beta\partial\beta'} \Big) \right)^{-1} \frac{\partial \log L(\beta^{(j)})}{\partial\beta},$$

where $\left(\dfrac{\partial^2 \log L(\beta^{(j)})}{\partial\beta\partial\beta'} \right)$ is replaced by $\mathrm{E}\left(\dfrac{\partial^2 \log L(\beta^{(j)})}{\partial\beta\partial\beta'} \right)$.

[End of Review]

In this case, we have the following iterative procedure:

$$\beta^{(j+1)} = \beta^{(j)} - \left(-\sum_{i=1}^{n} X_i' X_i \exp(X_i\beta^{(j)}) \right)^{-1} \left(-\sum_{i=1}^{n} X_i' \exp(X_i\beta^{(j)}) + \sum_{i=1}^{n} X_i' y_i \right).$$

The Newton-Raphson method is equivalent to the scoring method in this count model, because any random variable is not included in the expectation.