Example 6: Nested logit model:

(i) In the 1st step, choose YES or NO. Each probability is P_Y and $P_N = 1 - P_Y$.

(ii) Stop if NO is chosen in the 1st step. Go to the next if YES is chosen in the 1st step.

(iii) In the 2nd step, choose A or B if YES is chosen in the 1st step. Each probability is $P_{A|Y}$ and $P_{B|Y}$.

For simplicity, usually we assume the logistic distribution.

So, we call the nested logit model.

The probability that the ith individual chooses NO is:

$$P_{N,i} = \frac{1}{1 + \exp(X_i\beta)}.$$

The probability that the ith individual chooses YES and A is:

$$P_{A|Y,i}P_{Y,i} = P_{A|Y,i}(1 - P_{N,i}) = \frac{\exp(Z_i\alpha)}{1 + \exp(Z_i\alpha)} \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}.$$

The probability that the ith individual chooses YES and B is:

$$P_{B|Y,i}P_{Y,i} = (1 - P_{A|Y,i})(1 - P_{N,i}) = \frac{1}{1 + \exp(Z_i\alpha)} \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}.$$

In the 1st step, decide if the ith individual buys a car or not.

In the 2nd step, choose A or B.

X_i includes annual income, distance from the nearest station, and so on.

Z_i are speed, fuel-efficiency, car company, color, and so on.

The likelihood function is:

$$\begin{aligned} L(\alpha, \beta) &= \prod_{i=1}^{n} P_{N,i}^{I_{1i}} \Big(((1 - P_{N,i})P_{A|Y,i})^{I_{2i}} ((1 - P_{N,i})(1 - P_{A|Y,i}))^{1-I_{2i}} \Big)^{1-I_{1i}} \\ &= \prod_{i=1}^{n} P_{N,i}^{I_{1i}} (1 - P_{N,i})^{1-I_{1i}} \Big(P_{A|Y,i}^{I_{2i}} (1 - P_{A|Y,i})^{1-I_{2i}} \Big)^{1-I_{1i}}, \end{aligned}$$

where

$$I_{1i} = \begin{cases} 1, & \text{if the } i\text{th individual decides not to buy a car in the 1st step,} \\ 0, & \text{if the } i\text{th individual decides to buy a car in the 1st step,} \end{cases}$$

$$I_{2i} = \begin{cases} 1, & \text{if the } i\text{th individual chooses A in the 2nd step,} \\ 0, & \text{if the } i\text{th individual chooses B in the 2nd step,} \end{cases}$$

Remember that $\mathrm{E}(y_i) = F(X_i\beta^*)$, where $\beta^* = \dfrac{\beta}{\sigma}$.
Therefore, size of β^* does not mean anything.

The marginal effect is given by:

$$\frac{\partial \mathrm{E}(y_i)}{\partial X_i} = f(X_i\beta^*)\beta^*.$$

Thus, the marginal effect depends on the height of the density function $f(X_i\beta^*)$.

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:

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

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

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

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 $z=\dfrac{x-\mu}{\sigma}$. $\qquad x>a\implies z=\dfrac{x-\mu}{\sigma}>\dfrac{a-\mu}{\sigma}$.

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 $t=\frac{1}{2}z^2$. $\qquad z>\frac{a-\mu}{\sigma} \Longrightarrow t=\frac{1}{2}z^2>\frac{1}{2}(\frac{a-\mu}{\sigma})^2.$

[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 $y_i \le f(K_i, L_i)$, i.e., $u_i \le 0$.

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