Econometrics II TA Session #3

Ryo Sakamoto

Oct 27, 2021

Contents

- Discrete Choice Model
  - Example 1: Binary choice model
  - Example 2
  - Example 3
  - Example 4: Oredered probit and logit model
  - Example 5: Multinomial logit model
  - Example 6: Nested logit model

Example 1: Binary Choice Model

- Consider the regression model:

$$y_i^* = X_i\beta + u_i, \quad u_i \sim (0, \sigma^2), \quad i = 1, 2, \cdots, n,$$

where y_i^* is unobserved, but y_i is observed as 0 or 1, i.e.,

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

- E.g.) y_i^*: productivity of market work (continuous variable)

  y_i: whether an individual is employed or not (discrete variable)

- Note that we do not specify the distribution of u_i.

- Consider the probability that y_i takes 1, i.e.,

$$\begin{aligned}
\mathbb{P}(y_i = 1) &= \mathbb{P}(y_i^* > 0) \\
&= \mathbb{P}(u_i > -X_i\beta) \\
&= \mathbb{P}\left(\frac{u_i}{\sigma} > -X_i\frac{\beta}{\sigma}\right) \\
&= \mathbb{P}(u_i^* > -X_i\beta^*) \\
&= 1 - \mathbb{P}(u_i^* \leq -X_i\beta^*) \\
&= 1 - F(-X_i\beta^*) \\
&= F(X_i\beta^*),
\end{aligned}$$

where the last equality holds if the distribution of u_i^* is symmetric.

- The distribution function of u_i^* is $F(x) = \int_{-\infty}^{x} f(z)dz$.
  - If u_i^* follows standard normal distribution, we call Probit model.

$$F(x) = \int_{-\infty}^{x} \frac{1}{2\pi} \exp\left(-\frac{1}{2}z^2\right) dz$$

  - If u_i^* follows logistic distribution, we call Logit model.

$$F(x) = \frac{1}{1 + \exp(-x)}$$

- Since y_i is a binary variable, y_i follows Bernoulli distribution.
- Then, the density function of y_i is given by

$$
\begin{aligned}
f(y_i) &= \left[\mathbb{P}(y_i = 1)\right]^{y_i} \left[\mathbb{P}(y_i = 0)\right]^{1-y_i} \\
&= \left[F(X_i\beta^*)\right]^{y_i} \left[1 - F(X_i\beta^*)\right]^{1-y_i}.
\end{aligned}
$$

- Using this density function, we define the likelihood function:

$$
\begin{aligned}
L(\beta^*) &= f(y_1, \cdots, y_n) \\
&= \prod_{i=1}^{n} f(y_i) \\
&= \prod_{i=1}^{n} \left[F(X_i\beta^*)\right]^{y_i} \left[1 - F(X_i\beta^*)\right]^{1-y_i}.
\end{aligned}
$$

- The log-likelihood function is:

$$
\log L(\beta^*) = \sum_{i=1}^{n} \left[y_i \log F(X_i\beta^*) + (1 - y_i) \log[1 - F(X_i\beta^*)] \right].
$$

- The F.O.C. is:

$$\frac{\partial \log L(\beta^*)}{\partial \beta^*} = \sum_{i=1}^{n} \left(\frac{y_i X_i' f(X_i\beta^*)}{F(X_i\beta^*)} - \frac{(1-y_i) X_i' f(X_i\beta^*)}{1-F(X_i\beta^*)} \right)$$
$$= \sum_{i=1}^{n} \frac{X_i' f_i (y_i - F_i)}{F_i(1-F_i)} = 0,$$

where $f_i = f(X_i\beta^*)$ and $F_i = F(X_i\beta^*)$.

- The S.O.C. is:

$$\frac{\partial^2 \log L(\beta^*)}{\partial \beta^* \partial \beta^{*\prime}} = \sum_{i=1}^{n} \frac{X_i' \frac{\partial f_i}{\partial \beta^*}(y_i - F_i)}{F_i(1-F_i)} + \sum_{i=1}^{n} \frac{X_i' f_i \frac{\partial (f_i - F_i)}{\partial \beta^*}}{F_i(1-F_i)} + \sum_{i=1}^{n} X_i' f_i (y_i - F_i) \frac{\partial [F_i(1-F_i)]^{-1}}{\partial \beta^*}$$

$$= \sum_{i=1}^{n} \frac{X_i' X_i f_i' (y_i - F_i)}{F_i(1-F_i)} - \sum_{i=1}^{n} \frac{X_i' X_i f_i^2}{F_i(1-F_i)} + \sum_{i=1}^{n} X_i' f_i (y_i - F_i) \frac{X_i f_i (1-2F_i)}{[F_i(1-F_i)]^2}$$

is a negative definite matrix.

- For maximization, the method of scoring is:

$$\beta^{*(j+1)} = \beta^{*(j)} + \left[- \mathbb{E}\left(\frac{\partial^2 \log L(\beta^{*(j)})}{\partial \beta^* \partial \beta^{*\prime}} \right) \right]^{-1} \frac{\partial \log L(\beta^{*(j)})}{\partial \beta^*}$$

$$= \beta^{*(j)} + \left[\sum_{i=1}^{n} \frac{X_i' X_i \left(f_i^{(j)}\right)^2}{F_i^{(j)}\left(1 - F_i^{(j)}\right)} \right]^{-1} \sum_{i=1}^{n} \frac{X_i' f_i^{(j)} \left(y_i - F_i^{(j)}\right)}{F_i^{(j)}\left(1 - F_i^{(j)}\right)},$$

where $F_i^{(j)} = F(X_i \beta^{*(j)})$ and $f^{(j)} = f(X_i \beta^{*(j)})$.

- Note taht we use the following relationship:

$$\mathbb{E}[y_i] = \mathbb{P}(y_i = 1) = F_i(X_i \beta^*) = F_i.$$

- The Fisher information matrix is given by:

$$I(\beta^*) = -\mathbb{E}\left[\frac{\partial^2 \log L(\beta^*)}{\partial\beta^*\partial\beta^{*\prime}}\right] = \sum_{i=1}^{n} \frac{X_i' X_i f_i^2}{F_i(1-F_i)}.$$

- By the asymptotic normality,

$$\sqrt{n}(\hat{\beta}^* - \beta^*) \xrightarrow{d} \mathcal{N}\left(0, \lim_{n\to\infty}\left(-\frac{1}{n}\mathbb{E}\left[\frac{\partial^2 \log L(\beta^*)}{\partial\beta^*\partial\beta^{*\prime}}\right]^{-1}\right)\right),$$

where $\hat{\beta}^* := \lim_{j\to\infty}\beta^{*(j)}$ denotes the MLE of β^*.

Example 2

- Consider the two utility functions:

$$U_{1i} = X_i\beta_1 + \epsilon_{1i}, \tag{1}$$

$$U_{2i} = X_i\beta_2 + \epsilon_{2i}. \tag{2}$$

- We purchase a good when $U_{1i} > U_{2i}$ and do not purchase otherwise.
- y_i takes 1 if we purchase the good and takes 0 otherwise.
- We can observe y_i, but can NOT observe $U_{ji},\ j \in \{1, 2\}$.

- Taking a difference between (1) and (2), we have

$$U_{1i} - U_{2i} = X_i(\beta_1 - \beta_2) + (\epsilon_{1i} - \epsilon_{2i})$$
$$\iff \quad U_i^* = X_i\beta^* + \epsilon^*,$$

where $U_i^* := U_{1i} - U_{2i}$, $\beta^* := \beta_1 - \beta_2$ and $\epsilon^* := \epsilon_{1i} - \epsilon_{2i}$.

- Then, we have the following relationship:

$$y_i = \begin{cases} 1 & \text{if } U_i^* > 0, \\ 0 & \text{if } U_i^* \leq 0, \end{cases}$$

which is the same situation as Example 1 and the assumption that ϵ_i^* follows a symmetric distribution is necessary.

Example 3

- Consider the questionnaire:

$$y_i = \begin{cases} 1, & \text{if the } i\text{th person answers YES,} \\ 0, & \text{if the } i\text{th person answers NO.} \end{cases}$$

- Consider the following linear regression model:

$$y_i = X_i\beta + u_i.$$

- For instance, the question is "Do you have a car?" and X_i includes income, living place, and family size, etc.

- When $\mathbb{E}[u_i] = 0$,

$$\mathbb{E}[y_i] = X_i\beta.$$

- Because of the linear function, $X_i\beta$ takes the value from $-\infty$ to ∞.
- For instance, if an individual i has a high income, lives in the countryside and has many children, $X_i\beta$ takes a value that is greater than 1.
- However, since $\mathbb{E}[y_i]$ means the probability that an individual i has a car, $\mathbb{E}[y_i]$ must be in $[0, 1]$.

- Alternatively, consider the following model:

$$y_i = \mathbb{P}(y_i = 1) + u_i,$$

where u_i is a discrete type of random variable, i.e.,

$$u_i = \begin{cases} 1 - \mathbb{P}(y_i = 1) & \text{with prob. } \mathbb{P}(y_i = 1), \\ -\mathbb{P}(y_i = 1) & \text{with prob. } 1 - \mathbb{P}(y_i = 1) = \mathbb{P}(y_i = 0). \end{cases}$$

- Consider that $\mathbb{P}(y_i = 1)$ is connected with the distribution function $F(X_i\beta)$ as follows:

$$\mathbb{P}(y_i = 1) = F(X_i\beta).$$

- Assuming that $F(\cdot)$ is normal distribution or logistic distribution results in probit model or logit model, respectively.

Example 4: Ordered probit or logit model

- Consider the regression model:

$$y_i^* = X_i\beta + u_i, \quad u_i \sim (0,1), \quad i = 1, \cdots, n,$$

where y_i^* is unobserved, but y_i is observed as $1, 2, \cdots, m$, i.e.,

$$y_i = \begin{cases} 1, & \text{if } -\infty < y_i^* \le a_1, \\ 2, & \text{if } a_1 < y_i^* \le a_2, \\ \vdots & \\ m, & \text{if } a_{m-1} < y_i^* \le \infty, \end{cases}$$

where $a_1, \cdots, a_{m-1}$ are assumed to be known.

- For instance, y_i^* is hours worked per week and y_i is a discrete variable such that

$$y_i = \begin{cases} 1, & \text{if } y_i^* \leq 15, \\ 2, & \text{if } 15 < y_i^* \leq 20, \\ \vdots & \\ 9, & \text{if } 50 < y_i^*. \end{cases}$$

- Consider the probability that y_i takes $1, \cdots, m$.

$$\begin{aligned}
\mathbb{P}(y_i = 1) &= \mathbb{P}(y_i^* \leq a_1) \\
&= \mathbb{P}(u_i < a_1 - X_i\beta) \\
&= F(a_1 - X_i\beta). \\
\mathbb{P}(y_i = 2) &= \mathbb{P}(a_1 < y_i^* \leq a_2) \\
&= \mathbb{P}(a_1 - X_i\beta < u_i \leq a_2 - X_i\beta) \\
&= F(a_2 - X_i\beta) - F(a_1 - X_i\beta).
\end{aligned}$$

$$\begin{aligned}\mathbb{P}(y_i = m) &= \mathbb{P}(a_{m-1} < y_i^*) \\ &= \mathbb{P}(a_{m-1} - X_i\beta < u_i) \\ &= 1 - F(a_{m-1} - X_i\beta).\end{aligned}$$

- We have

$$\mathbb{P}(y_i = j) = F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta),\ \forall j \in \{0, 1, \cdots, m\}$$

where $a_0 = -\infty$ and $a_m = \infty$.

- Define the following indicator functions:

$$I_{ij} = \begin{cases} 1, & \text{if } y_i = j, \\ 0, & \text{otherwise.} \end{cases}$$

- The likelihood function is:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} \left[F(a_1 - X_i\beta)\right]^{I_{i1}} \left[F(a_2 - X_i\beta) - F(a_1 - X_i\beta)\right]^{I_{i2}} \cdots \left[1 - F(a_{m-1} - X_i\beta)\right]^{I_{im}} \\ &= \prod_{i=1}^{n}\prod_{j=1}^{m} \left[F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta)\right]^{I_{ij}}. \end{aligned}$$

- The log-likelihood function is:

$$\log L(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{ij} \log \Big[F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta) \Big].$$

- The first derivative of $\log L(\beta)$ w.r.t. β is:

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{-I_{ij} X_i' \Big[f(a_j - X_i\beta) - f(a_{j-1} - X_i\beta) \Big]}{F(a_j - X_i\beta) - F(a_{j-1} - X_i\beta)} = 0.$$

- Usually, normal distribution or logistic distribution is chosen for $F(\cdot)$.

Example 5: Multinomial logit model

- The ith individual has $m+1$ choices, i.e., $j = 0, 1, \cdots, m$.

$$\mathbb{P}(y_i = j) = \frac{\exp(X_i\beta_j)}{\sum_{j=0}^{m}\exp(X_i\beta_j)}$$
$$= \frac{\exp(X_i\beta_j)}{\exp(X_i\beta_0) + \exp(X_i\beta_1) + \cdots + \exp(X_i\beta_m)}$$
$$= \frac{\exp(X_i\beta_j)}{1 + \exp(X_i\beta_1) + \cdots + \exp(X_i\beta_m)} =: P_{ij},$$

where $\beta_0 = 0$.

- Different from the ordered probit or logit model, the order does not matter.
- For instance, X_i is IQ and y_i indicates occupations: y_i takes 1 if i is a cook, takes 2 if i is a professor and takes 3 if i is an artist.
- In using the multinomial logit model, a choice is set to be a comparison.
- Therefore, if we set being a cook as a comparison, we interpret the estimation results as follows:
  - one unit of increase in IQ increases the probability of being a professor relative to being a cook by A%;
  - one unit of increase in IQ increases the probability of being an artist relative to being a cook by B%.

- Note that

$$P_{i0} = \frac{1}{1 + \exp(X_i\beta_1) + \cdots + \exp(X_i\beta_m)}$$

- Then, we have

$$\frac{P_{ij}}{P_{i0}} = \exp(X_i\beta_j) \iff \log\frac{P_{ij}}{P_{i0}} = X_i\beta_j.$$

- The log-likelihood function is:

$$\log L(\beta_1, \cdots, \beta_m) = \sum_{i=1}^{n}\sum_{j=1}^{m} d_{ij}\log P_{ij},$$

where $d_{ij} = 1$ when the ith individual chooses jth choice, and $d_{ij} = 0$ otherwise.

Example 6: Nested logit model

- Consider the following 2 steps:
  1. Choose YES or NO with probability P_Y and $P_N = 1 - P_Y$, respectively. Go to the next step only if YES is chosen.
  2. Choose A or B with probability $P_{A|Y}$ and P_{B_Y}, respectively.
- For instance, the individual decides whether or not to buy a car in the first step and chooses Audi or BMW.
- Assume the logistic distribution.

- 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:

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

- X_i means variables which affect the decision making on whether to buy a car:
  - annual income
  - distance from the nearest station
- Z_i means variables which characterize a car:
  - speed
  - fuel-efficiency
  - car company

- The likelihood function is:

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

where

- I_{1i} takes 1 if ith individual decides not to buy a car in the first step and takes 0 otherwise;
- I_{2i} takes 1 if ith individual chooses A in the second step and takes 0 otherwise.

- Let individual $i \in \mathcal{N}$ decide not to buy a car, $i \in \mathcal{A}$ choose A and $i \in \mathcal{B}$ choose B.
- Then, the likelihood function becomes:

$$\begin{aligned} L(\alpha, \beta) &= \prod_{i=1}^{n} P_{N,i}^{I_{1i}} (1 - P_{N,i})^{1-I_{1i}} \left[P_{A|Y,i}^{I_{2i}} (1 - P_{A|Y,i})^{1-I_{2i}} \right]^{1-I_{1i}} \\ &= \prod_{i \in \mathcal{N}} P_{N,i} \times \prod_{i \in \mathcal{A}} (1 - P_{N,i}) P_{A|Y,i} \times \prod_{i \in \mathcal{B}} (1 - P_{N,i})(1 - P_{A|Y,i}). \end{aligned}$$

- The log-likelihood function is:

$$\log L(\alpha, \beta) = \sum_{i \in \mathcal{N}} P_{N,i} + \sum_{i \in \mathcal{A}} (1 - P_{N,i}) P_{A|Y,i} + \sum_{i \in \mathcal{B}} (1 - P_{N,i})(1 - P_{A|Y,i}).$$

- Substituting the expressions above into the log-likelihood yields:

$$\begin{aligned}\log L(\alpha,\beta) &= \sum_{i\in\mathcal{N}} P_{N,i} + \sum_{i\in\mathcal{A}} (1-P_{N,i})P_{A|Y,i} + \sum_{i\in\mathcal{B}} (1-P_{N,i})(1-P_{A|Y,i}) \\ &= \sum_{i\in\mathcal{N}} \frac{1}{1+\exp(X_i\beta)} + \sum_{i\in\mathcal{A}} \frac{\exp(Z_i\alpha)}{1+\exp(Z_i\alpha)} \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)} \\ &+ \sum_{i\in\mathcal{B}} \frac{1}{1+\exp(Z_i\alpha)} \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)}.\end{aligned}$$

- Using this, we can consider the F.O.C. and S.O.C. w.r.t. α and β.