

Characteristics of Discrete Dependent Data

Categorical Data

- Binary-choice Data
- Multi-choice Data
 - Unordered Data
 - Ordered Data
 - Sequential Data

Non-categorical Data

- Counted Number

Characteristics of Discrete Dependent Data

Categorical Data

- Binary-choice Data

1. Binary Data → 1. Logit & Probit

- Multi-choice Data

2. Multinomial Data → 2. MN Logit & Probit

3. Multivariate Data → 3. Multivariate Probit

4. Ordered Data → 4. OLogit & Probit

Non-categorical Data

5. Count Number → 5. Poisson Model

Multinomial Data

Dependent variable has a finite number of possible outcomes.

Individuals can choose among more than two choices.

Individual can choose only one choice.

Multinomial Model

Let m be number of alternatives.

Response (dependent) variable $y_i = j$ is nominal variable, where $j = 1, 2, \dots, m$.

Then, model for stochastic utilities is given by:

$$U_i^j = u_{ij} + \varepsilon_{ij} = x_{ij}'\beta + \varepsilon_{ij}$$

where:

x_{ij} is vector of explanatory variables that apply for individual i and alternative j .

$x_{ij} - x_{ih}$ measures the differences between alternative j and alternative h – e.g. differences in interest rates between strategy j and h .

Choice Model and Log-likelihood

Assume maximize utility.

$$p_{ij} = \Pr[y_i = j] = \Pr[u_{ij} + \varepsilon_{ij} > u_{ih} + \varepsilon_{ih} \text{ for all } h \neq j]$$

Assume individuals make *independent* choice, the log-likelihood function can be stated as:

$$\log L = \sum_{i=1}^n \sum_{j=1}^m y_{ij} \log p_{ij} = \sum_{i=1}^n \log p_{iy_i}$$

where:

$y_{ij} = 1$ if $y_i = j$ and $y_{ij} = 0$ otherwise.

Multinomial Probit Model

Assume joint normal distribution for the model.

$$\begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{im} \end{pmatrix} \sim NID(0, V)$$

The model is multinomial probit model.

Multinomial Logit Model

Assume that the error terms are independently and identically distributed.

Multinomial Logit:
$$p_{ij} = \frac{e^{x'_{ij}\beta}}{\sum_{h=1}^m e^{x'_{ih}\beta}}$$

Marginal Effects

Marginal effects of the explanatory variables on the choice probabilities:

Multinomial Logit:

$$\frac{\partial \Pr_{CL}[y_i = j]}{\partial x_{ij}} = p_{ij} (1 - p_{ij}) \beta$$

Ordered Response Model

In case that choices are ordered, ordered probit or logit model can be applied.

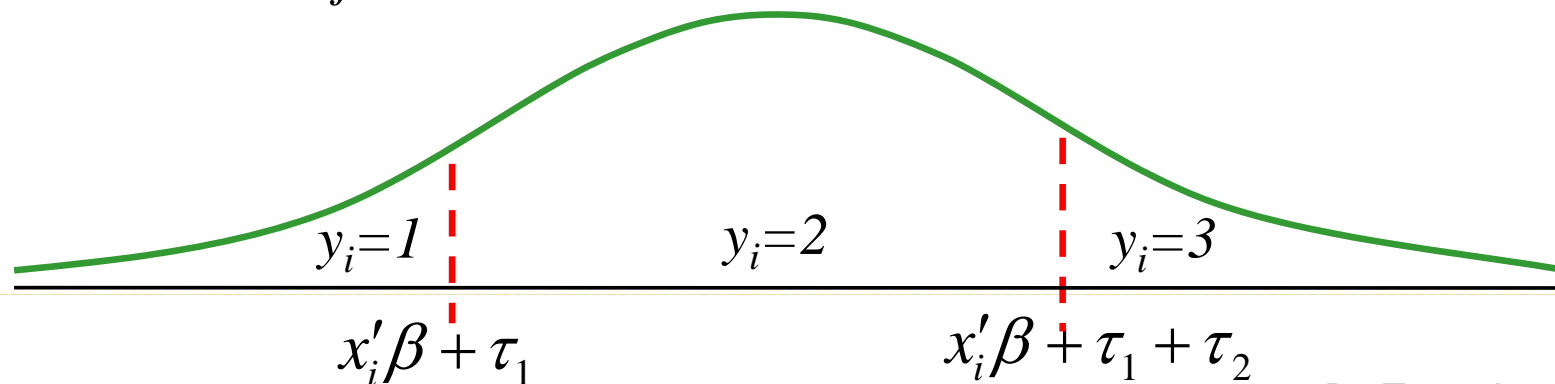
$$y_i^* = x_i' \beta + \varepsilon_i, \quad E(\varepsilon_i) = 0$$

$$y_i = 1 \quad \text{if } -\infty < y_i^* \leq \tau_1 + x_i' \beta,$$

$$y_i = j \quad \text{if } \sum_{k=1}^{j-1} \tau_k + x_i' \beta < y_i^* \leq \sum_{k=1}^j \tau_k + x_i' \beta, \quad j = 2, \dots, m-1,$$

$$y_i = m \quad \text{if } \sum_{k=1}^{m-1} \tau_k + x_i' \beta < y_i^* \leq \infty$$

where $\tau_j =$ Threshold value, $j=1,2,\dots, m$



Multivariate Probit Model

Dependent variable has more than two choices.

Individual can choose more than one choice.

The model assumes multivariate distribution.

$$y_1^* = \beta_1' x_1 + \varepsilon_1, \quad y_1 = 1 \text{ if } y_1^* > 0, \quad 0 \text{ otherwise}$$
$$y_2^* = \beta_2' x_2 + \varepsilon_2, \quad y_2 = 1 \text{ if } y_2^* > 0, \quad 0 \text{ otherwise}$$

$$E[\varepsilon_1 | x_1, x_2] = E[\varepsilon_2 | x_1, x_2] = 0$$

$$\text{Var}[\varepsilon_1 | x_1, x_2] = \text{Var}[\varepsilon_2 | x_1, x_2] = 1$$

$$\text{Cov}[\varepsilon_1, \varepsilon_2 | x_1, x_2] = \rho$$