Multinomial Probit and Logit Models
Conditional Logit Model
Mixed Logit Model

Ani Katchova

Multinomial, Conditional, and Mixed Models Overview

- Multinomial outcome dependent variable (in wide and long form of data sets)
- Independent variables (alternative-invariant or alternative-variant)
- Multinomial logit model (coefficients, marginal effects, IIA) and multinomial probit model
- Conditional logit model (coefficients, marginal effects)
- Mixed logit model

Multinomial, Conditional and Mixed Models

Multinomial outcome examples

- The type of insurance contract that an individual selects.
- The product that an individual selects (say type of cereal).
- Occupational choice by an individual (business, academic, non-profit organization).
- The choice of fishing mode (beach, pier, private boat, charter boat).

Multinomial outcome dependent variable

- The dependent variable y is a categorical, unordered variable.
- An individual may select only one alternative.
- The choices/categories are called alternatives and are coded as j =1, 2, …, m.
- The numbers are only codes and their magnitude cannot be interpreted (use frequency for each category instead of means to summarize the dependent variable).
- The data are usually recorded in two formats: a wide format and a long format.
- When using the wide format, the data for each individual i is recorded on one row. The dependent variable is:

$$y = j$$

- When using the long format, the data for each individual *i* is recorded on *j* rows, where *j* is the number of alternatives. The dependent variable is:

$$y_j = \begin{cases} 1 \text{ if } y = j \\ 0 \text{ if } y \neq j \end{cases}$$

- Therefore, $y_j = 1$ if the alternative *j* is the observed outcome and the remaining $y_k = 0$. For each observation only one of $y_1, y_2, \ldots, y_m$ will be non-zero.

Example for multinomial data in wide form

Person ID (*i*)	Dependent variable (y)	Codes for y	w_i (income)	x_{i1} (price of alternative 1)	x_{i2} (price of alternative 2)
1	apple juice (alternative 1)	y=1	40,000	2.5	1.5
2	orange juice (alternative 2)	y=2	38,000	2.7	1.7
3	orange juice (alternative 2)	y=2	50,000	2.9	1.6

Example for multinomial data in long form

Person ID (i)	Dependent variable (y_j)	Codes for y_j	w_i (income)	x_{ij} (price)
1	apple juice (alternative 1)	$y_1 = 1$	40,000	2.5
1	orange juice (alternative 2)	$y_2 = 0$	40,000	1.5
2	apple juice (alternative 1)	$y_1 = 0$	38,000	2.7
2	orange juice (alternative 2)	$y_2 = 1$	38,000	1.7
3	apple juice (alternative 1)	$y_1 = 0$	50,000	2.9
3	orange juice (alternative 2)	$y_2 = 1$	50,000	1.6

- The multinomial density for one observation is defined as:

$$f(y) = p_1^{y_1} \times \ldots \times p_m^{y_m} = \prod_{j=1}^{m} p_j^{y_j}$$

- The probability that individual i chooses the jth alternative is:

$$p_{ij} = \mathrm{pr}[y_i = j] = \mathrm{F}_j(\mathbf{x}_i, \beta)$$

- The functional form of F_j should be selected so that the probabilities lie between 0 and 1 and sum over j to one. Different functional forms of F_j lead to multinomial, conditional, mixed, and ordered logit and probit models.

Independent variables

- Two types of independent variables.
- *Alternative-invariant or case-specific regressors* –the regressors w_i vary over the individual *i* but do not vary over the alternative *j*.
    - Income, age, and education are different for each individual but they do not vary based on the type of a product that the individual selects.
    - Used in the multinomial logit model.
- *Alternative-variant or alternative-specific regressors* – the regressors x_{ij} vary over the individual *i* and the alternative *j*.
    - Prices for products vary for each product and individuals may also pay different prices.
    - Salaries for occupation may be different between occupations and also for each individual.
    - Used in the conditional and mixed logit models.

Multinomial logit model

- The multinomial logit model is used with alternative-invariant regressors.
- The probability that individual *i* will select alternative *j* is:

$$p_{ij} = p(y_i = j) = \frac{\exp(\mathbf{w}_i'\gamma_j)}{\sum_{\mathrm{k=1}}^{\mathrm{m}} \exp(\mathbf{w}_i'\gamma_k)}$$

- This model is a generalization of the binary logit model.
- The probabilities for choosing each alternative sum up to 1, $\sum_{\mathrm{j=1}}^{\mathrm{m}} p_{ij} = 1$
- One set of coefficients needs to be normalized to zero to estimate the models (usually $\gamma_1 = 0$), so there are (j-1) sets of coefficients estimated. The coefficients of other alternatives are interpreted in reference to the base outcome.
- Coefficient interpretation for alternative *j*: in comparison to the base alternative, an increase in the independent variable makes the selection of alternative *j* more or less likely.

Marginal effects

- The marginal effect of an increase of a regressor on the probability of selecting alternative j is:
$$\partial p_{ij}/\partial \mathbf{w}_i = p_{ij}(\gamma_j - \overline{\gamma_\imath})$$
- The marginal effects do not necessarily correspond in sign to the coefficients (unlike the binary logit or probit model).
- There are (j-1) sets of coefficients because one set is normalized to zero, but there are j sets of marginal effects.
- Depending on which alternative we select as a base category, the coefficients will be different (in reference to the base category) but the marginal effects will be the same regardless of the base category.
- The marginal effects of each variable on the different alternatives sum up to zero.
- Marginal effects interpretation: each unit increase in the independent variable increases/decreases the probability of selecting alternative j by the marginal effect expressed as a percent.

Independence from Irrelevant Alternatives (IIA) property

- The odds ratios in the multinomial logit models are independent of other alternatives. For choices j and k, the odds ratio only depends on the coefficients for choices j and k.
- Odds ratio: $p_{ij}/p_{ik} = \exp\left(\mathbf{w}_i'(\gamma_j - \gamma_k)\right)$
- This weakness of the multinomial model is known as the red bus-blue bus problem. If the choice is between a car and a blue bus, according to the model the introduction of a red bus will not change the probabilities.

Multinomial probit model

- The multinomial probit model is similar to multinomial logit model, just like the binary probit model is similar to the binary logit model.
- The difference is that it uses the standard normal cdf.
- The probability that observation *i* will select alternative j is:

$$p_{ij} = p(y_i = j) = \Phi\,(\mathbf{x}_{ij}'\beta)$$

- It takes longer for a probit model to obtain results.
- The coefficients are different by a scale factor from the logit model.
- The marginal effects will be similar.

Conditional logit model

- The conditional logit model is used with alternative-invariant and alternative-variant regressors.
- The probability that observation *i* will choose alternative *j* is:

$$p_{ij} = p(y_i = j) = \frac{\exp\left(\mathbf{x}_{ij}'\beta + \mathbf{w}_i'\gamma_j\right)}{\sum_{\mathrm{k}=1}^{\mathrm{m}} \exp\left(\mathbf{x}_{ik}'\beta + \mathbf{w}_i'\gamma_k\right)}$$

  where $\mathbf{x}_{ij}$ are alternative-specific regressors and $\mathbf{w}_i$ are case-specific regressors.
- The conditional logit model has (*j*-1) sets of coefficients (γ_j) (with one set being normalized to zero) for the case-specific regressors and only one set of coefficients (β) for the alternative-specific regressors.
- The probabilities for choosing each alternative sum up to 1.
- Coefficients for the alternative-invariant regressors γ_j (similar treatment as the multinomial logit model).
  - One set of coefficients for the alternative-invariant regressors is normalized to zero (say $\gamma_1 = 0$), this is the base outcome. The rest of coefficients are interpreted in relation to this base category.

- There are (j-1) sets of coefficients (corresponding to the number of alternatives minus 1 for the base).
- Coefficient interpretation for alternative j: in comparison to the base alternative, an increase in the independent variable makes the selection of alternative j more or less likely.

- Coefficients for the alternative-specific regressors (β).
  - No normalization is needed.
  - One set of coefficients across all alternatives.
  - Coefficient interpretation: an increase in the price of one alternative decreases the probability of choosing that alternative and increases the probability of choosing other alternatives.

Marginal effects

- The marginal effect of an increase of a regressor on the probability of selecting alternative *j* is:

$$\partial p_{ij}/\partial \mathbf{x}_{ik} = p_{ij}(\delta_{ijk} - p_{ik})\beta$$

  where $\delta_{ijk} = 1$ if j=k and 0 otherwise.

- There are *j* sets of marginal effects for both the alternative-specific and case-specific regressors.
- For each alternative-specific variable $\mathbf{x}_{ij}$, there are *j*x*j* sets of marginal effects.
- The marginal effects of each variable on the different alternatives sum up to zero.
- Marginal effects interpretation: each unit increase in the independent variable increases the probability of selecting the *k*th alternative and decreases the probability of the other alternatives, by the marginal effect expressed as a percent.

Mixed logit model

The mixed logit model (also called random parameters logit model) specifies the utility to the *i*th individual for the *j*th alternative to be:

$$U_{ij} = \mathbf{x}'_{ij}\beta_i + \mathbf{w}'_i\gamma_{ji} + e_{ij} = \mathbf{x}'_{ij}\beta + \mathbf{w}'_i\gamma_j + \mathbf{x}'_{ij}\upsilon_i + \mathbf{w}'_i\delta_{ji} + e_{ij}$$

where e_{ij} are iid extreme value (similar to the errors in the conditional logit model).

- The mixed logit model allows for the parameters β_i to be random. A common assumption is that $\beta_i = \beta + \upsilon_i$ where $\upsilon_i \sim \mathrm{N}[0, \Sigma_\beta]$ and $\gamma_{ji} = \gamma_j + \delta_{ji}$ where $\delta_{ji} \sim \mathrm{N}[0, \Sigma_{\gamma i}]$.
- The introduction of the random parameters has the attractive property of inducing correlation across alternatives. The combined error $\mathbf{x}'_{ij}\upsilon_i + \mathbf{w}'_i\delta_{ji} + e_{ij}$ is now correlated across alternatives, say $\mathrm{Cov}[\upsilon_{ij}, \upsilon_{ij}] = \mathbf{x}'_{ij}\Sigma_\beta\mathbf{x}_{ik}$.
- The probability that individual *i* selects alternative *j* represents a mixed logit model:

$$p_{ij} = p(y_i = j) = \frac{\exp\left(\mathbf{x}'_{ij}\beta + \mathbf{w}'_i\gamma_j + \mathbf{x}'_{ij}\upsilon_i + \mathbf{w}'_i\delta_{ji}\right)}{\sum_{\mathrm{k=1}}^{\mathrm{m}} \exp\left(\mathbf{x}'_{ik}\beta + \mathbf{w}'_i\gamma_k + \mathbf{x}'_{ik}\upsilon_i + \mathbf{w}'_i\delta_{ki}\right)}$$

- The mixed logit model relaxes the IIA assumption by allowing parameters in the conditional logit model to be normally (or log-normally) distributed.
- When estimating the mixed logit model, the researcher needs to specify which parameters will be estimated as random. If a parameter is random, this implies that effect of a particular regressor on the chosen alternative varies across the individuals.
- The mixed logit model produce random parameters coefficients for both the regressor (x_i) and the standard deviation of the regressor ($sd(x_i)$).
- Coefficient interpretation for the regressors (x_i): when the independent variable increases, the consumers are more or less likely to choose this alternative.
- Coefficient interpretation on the standard deviation of a regressor ($sd(x_i)$): there is a heterogeneity across individuals with respect to the effect of the independent variable on the alternative chosen.