Microeconometrics
Multinomial Choices

Pedro Portugal

Nova School of Business and Economics

Carcavelos Spring 2025

Basic Framework and Notation

- $i = \{1, 2, ..., N\}$ denotes a set of decision makers
- $j = \{0, 1, 2, 3, ..., H\}$ denotes a *finite* set of *mutually exclusive* and *exhaustive* possible choices

$$U_{ij} = X_{ij}\beta_j + \varepsilon_{ij}$$

represents the utility of the decision maker i if the choice is j and is a function of:

- a systematic component $X_{ij}\beta_j$ where
  - X_{ij} is a row vector of observed characteristics of the decision maker and of the choices
  - β_j is a column vector of unknown parameters which may change across choices
- a random unobservable component ε_{ij}

Basic Framework and Notation

- Let Y_i denote the indicator function that denotes which option has been chosen by the decision maker

$$Y_i = j \quad \text{if } i \text{ chooses } j$$

- Decision makers are assumed to maximize utility. Therefore:

$$Y_i = j \text{ if } U_{ij} > U_{is} \quad \forall s \neq j \text{ in the choice set}$$

- Since we observe only the systematic component of utility, we cannot predict with certainty the choice of each decision maker. We can only try to assess the probability that each decision maker will choose each alternative

Multinomial Logit Model

$$
\begin{aligned}
P_{ij} &= Pr(Y_i = j) \\
&= Pr(U_{ij} > U_{is}, \forall s \neq j) \\
&= Pr(X_{ij}\beta_j + \varepsilon_{ij} > X_{is}\beta_s + \varepsilon_{is}, \forall s \neq j) \\
&= Pr(\varepsilon_{is} - \varepsilon_{ij} < X_{ij}\beta_j - X_{is}\beta_s, \forall s \neq j)
\end{aligned}
$$

- If each ε_{ij} is independently distributed according to the *extreme value* cumulative distribution
$$\exp(-e^{-\varepsilon_{ij}})$$
- Then, the probability that the alternative j is chosen is given by the logit distribution

$$P_{ij} = \frac{e^{X_{ij}\beta_j}}{\sum_{s=0}^{H} e^{X_{is}\beta_s}}$$

Independence from Irrelevant Alternatives

- Implies that the odds of two alternatives j and s do not depend on the other existing alternatives:

$$\frac{P_{ij}}{P_{is}} = \frac{e^{X_{ij}\beta_j}}{e^{X_{is}\beta_s}}$$

which depends only on j and s

- This property may not be desirable. Consider the following classic example:
  - Initially there are only 2 options: $j =$"car" and $s =$"red bus"
  - Suppose $\frac{P_{ij}}{P_{is}} = 1$
  - Then a new option is added: $t =$"blue bus"
  - Suppose that decision makers who choose a bus are indifferent with respect to the color. Then we would expect the model to predict: $P_{ij} = 0.5$ and $P_{is} = P_{it} = 0.25$
  - However the logit model would continue to imply $\frac{P_{ij}}{P_{is}} = 1$
  - In order for this to be compatible with $P_{is} = P_{it}$, the estimated probabilities must be $P_{ij} = P_{is} = P_{it} = 1/3$

Identification

Consider as an example the problem of a consumer i who has to choose between *Japanese* ($j = 0$) or *European* ($j = 1$) cars

- The vector of attributes X_{ij} includes:
  - factors Z_{ij} that change across both individuals and choices (e.g. the price or the number of dealers of each car in the city where i lives
  - factors W_i that change only across individuals (e.g. gender, age or income of the consumer)
  - a choice specific constants α_j capturing factors that change across choices but not across individuals
- The vector of parameters to be estimated is $\beta_j' = \{\alpha_j, \gamma, \delta\}$, which differs across choices because there is a different constant for each choice
- The parameters γ and δ are assumed to be identical across choices

Identification

Under these assumptions the probability of the *european* choice would be

$$P_{i1} = Pr(Y_i = 1) = \frac{e^{\alpha_1 + Z_{i1}\gamma + W_i\delta}}{e^{\alpha_0 + Z_{i0}\gamma + W_i\delta} + e^{\alpha_1 + Z_{i1}\gamma + W_i\delta}}$$
$$= \frac{1}{1 + e^{(\alpha_0 - \alpha_1) + (Z_{i0} - Z_{i1})\gamma}}$$

This example highlights some identification problems in the logit model:

- if δ is identical across choices, this model cannot identify the effect of the decision maker's attributes ($W_i\delta$ cancels out)
- the model cannot identify the choice-specific constants but only the difference between them $\alpha_0 - \alpha_1$
- the model can identify the effects γ of the choice-specific attributes also if they are identical across choices

Multinomial Logit Model

- Conventional name for a multiple choice problem in which the representative utility of each choice depends only on the attributes of the decision maker
$$U_{ij} = X_i \beta_j + \varepsilon_{ij}$$
- To achieve identification the attributes are allowed to have different effects on the utility of the different choices
- In this case, the probability of a choice is given by:
$$\begin{aligned} P_{ij} &= \frac{e^{X_i \beta_j}}{\sum_{s=0}^{H} e^{X_i \beta_s}} \\ &= \frac{1}{\sum_{s=0}^{H} e^{X_i(\beta_s - \beta_j)}} \end{aligned}$$
therefore only differences between parameters can be identified
- It is convenient to impose the normalization with respect to a reference choice, for example $j = 0$

Multinomial Logit Model

Taking $j = 0$ as the reference choice means to impose the normalization $\beta_0 = 0$, which implies $e^{X_i\beta_0} = 1$ and therefore

$$
\begin{aligned}
P_{ij} &= Pr(Y_i = j) \\
&= \frac{e^{X_i\beta_j}}{1 + \sum_{s=1}^{H} e^{X_i\beta_s}}
\end{aligned}
$$

$$
\begin{aligned}
P_{i0} &= Pr(Y_i = 0) \\
&= \frac{1}{1 + \sum_{s=1}^{H} e^{X_i\beta_s}}
\end{aligned}
$$

Maximum Likelihood Estimator

The log-likelihood function of the Multinomial logit model is

$$\ln(L) = \sum_{i=1}^{N} \sum_{j=0}^{H} d_{ij} \ln(P_{ij})$$

where $d_{ij} = 1$ if i chooses j

The first-order conditions for the maximization of the likelihood are

$$\frac{\partial \ln(L)}{\partial \beta_j} = \sum_{i=1}^{N} (d_{ij} - P_{ij}) X_i = 0$$

This is a system of $K \times H$ equations

Maximum Likelihood Estimator

The second derivatives matrix is composed by H^2 blocks each with dimension $K \times K$

The *main diagonal* blocks have the form:

$$\frac{\partial^2 \ln(L)}{\partial \beta_j \beta_j'} = -\sum_{i=1}^{N} P_{ij}(1 - P_{ij}) X_i' X_i$$

The *off main diagonal* blocks (for $j \neq s$) have the form:

$$\frac{\partial^2 \ln(L)}{\partial \beta_j \beta_s'} = \sum_{i=1}^{N} P_{ij} P_{is} X_i' X_i$$

Interpretation

The parameters β_j should be interpreted carefully.

$$\ln \frac{P_{ij}}{P_{i0}} = X_i \beta_j$$

- The coefficient β_j measures the impact of the attributes X_i on the log-odds that the decision maker chooses j instead of s

Also,

$$\ln \frac{P_{ij}}{P_{is}} = X_i(\beta_j - \beta_s)$$

- The *difference* between the coefficients β_j and β_s measure the impact of the attributes X_i on the log-odds that the decision maker chooses j instead of s

Interpretation

The marginal effects of the individual attributes X_i on the probability of a choice $Y_i = j$ are even more difficult to interpret.

$$\gamma_j = \frac{\partial P_j}{\partial X_i} = P_j\Big(\beta_j - \sum_{s=0}^{H} P_s \beta_s\Big) = P_j(\beta_j - \bar{\beta})$$

- The effect of X_i on P_j (generic probability of a choice j) depends on the parameters concerning all the choices, not just on the parameters concerning choice j

Interpretation

The problem is that when X_i changes the probabilities of all the choices are contemporaneously affected. Consider the car example with 3 choices: "Japanese" ($j = 0$), "European" ($j = 1$) and "American" ($j = 2$)

- Suppose X_i is the age of the buyer and that $\beta_2 > \beta_1 > 0$
- Implies that older workers tend to buy more European and more American cars than Japanese cars. Moreover, older workers tend to buy more American cars than European cars
- However, if β_2 is much larger than β_1 it may happen that the probability of a European choice decreases, when age X_i increases

The marginal effects are a function of the explanatory factors X (since $P_j = \frac{e^{X\beta_j}}{\sum_{s=0}^{H} e^{X\beta_s}}$), and therefore have to be computed at some reference value of X (the mean, the median, a particular i,...)

Effects on Odds Ratio

As in the binary case, results can be expressed in the form of *odds ratios*, or exponentiated form

- The odds of a choice j instead of 0, given X_i, are:

$$\Omega(Y_i = j; Y_i = 0|X) = \frac{P_{ij}}{P_{i0}} = e^{X_i \beta_j}$$

- Given two realizations of X_i, X_1 and X_0, we can define the odds ratio

$$\frac{\Omega(Y_i = j; Y_i = 0|X_1)}{\Omega(Y_i = j; Y_i = 0|X_0)} = e^{(X_1 - X_0)\beta_j}$$

- This statistic tells us how the odds of observing $Y = j$ instead of $Y = 0$ change when X_i changes from X_0 to X_1

The Conditional Logit Model

Conventional name for a multiple choice problem in which the representative utility of each choice depends on choice specific attributes:

$$U_{ij} = X_{ij}\beta + \varepsilon_{it}$$

The probability of a choice model would be:

$$P_j = \frac{e^{X_{ij}\beta}}{\sum_{s=0}^{H} e^{X_{is}\beta}}$$

and in this case the coefficients β are identified even if they are identical across choices

Marginal Effects

Consider the car example and suppose that X_{ij} is the number of dealers in the city of each buyer, for each type of car

The marginal effect of an increase in the number of dealers of car j on the probability that car j is bought is:

$$\gamma_{ij} = \frac{\partial P_j}{\partial X_{ij}} = P_j(1 - P_j)\beta_{dealer}$$

The marginal effect of an increase in the number of dealers of car s on the probability that car j is bought is:

$$\gamma_{is} = \frac{\partial P_j}{\partial X_{is}} = -P_j P_s \beta_{dealer}$$

The usual odds ratio (exponentiated) representation of coefficients is also possible

Maximum Likelihood Estimation

The log-likelihood function of the Conditional Logit Model is

$$\ln(L) = \sum_{i=1}^{N} \sum_{j=0}^{H} d_{ij} \ln(P_{ij})$$

where $d_{ij} = 1$ if i chooses j

The first-order conditions for the maximization of the likelihood are:

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{N} \sum_{j=0}^{H} d_{ij}(X_{ij} - \bar{X}_i) = 0$$

where $\bar{X}_i = \sum_{j=0}^{H} P_{ij} X_{ij}$

The second derivatives matrix is:

$$\frac{\partial^2 \ln L}{\partial \beta \beta'} = \sum_{i=1}^{N} \sum_{j=0}^{H} P_{ij}(X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$$

A Test for the IIA Hypothesis

Hausman and McFadden (1984) suggest that if a subset of the choice set is irrelevant, omitting it from the model should not change the parameter estimates systematically

Consider a choice set $A = \{B, C\}$ where B and C are subsets of A. We want to test whether the presence of the choices in C are irrelevant for the odds between the choices in B

The statistic for the "Hausman's specification test" is

$$HM = (\widehat{\beta}_B - \widehat{\beta}_A)'[\widehat{V}_B - \widehat{V}_A]^{-1}(\widehat{\beta}_B - \widehat{\beta}_A) \sim \chi^2(K)$$

where

- $\widehat{\beta}_B$ and $\widehat{\beta}_A$ are the ML estimates of the parameters of the restricted and unrestricted models
- $\widehat{V}_B$ and $\widehat{V}_A$ are the ML estimates of the asymptotic covariance matrices of the restricted and unrestricted models
- Both estimates are consistent under the null and $\widehat{\beta}_A$ is more efficient

Multinomial Probit Model

Alternatives to the Conditional Logit Model: the Multinomial Probit Model

- Assume three alternative cases:

$$Y_{1i}^* = \mathbf{x}_{1i}\beta + u_{1i}$$
$$Y_{2i}^* = \mathbf{x}_{2i}\beta + u_{2i}$$
$$Y_{0i}^* = 0$$

where the unobservable components follow the joint normal process

$$\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right)$$

- Then

$$P[y_i = 1|\mathbf{x}_i] = P[y_{1i}^* > 0, y_{1i}^* > y_{2i}^*|\mathbf{x}_{1i}, \mathbf{x}_{2i}]$$
$$= \int_{-\mathbf{x}_{1i}\beta}^{+\infty} \int_{-\infty}^{(\mathbf{x}_{1i}-\mathbf{x}_{2i})\beta+u_1} f(u_1, u_2) du_2 du_1$$

Estimation of Multinomial Probit

The log-likelihood function for J possible choices is:

$$\ln L_N(\beta, \Omega) = \sum_{i=1}^{N} \sum_{j=0}^{J} \mathbf{1}(y_i = j) P(y_i = j | \mathbf{x}_i, \beta, \Omega)$$

where Ω is the variance-covariance matrix

For a large number of options, the solution is to use simulation

- **Simulated Maximum Likelihood**
  - Draw R random vectors y^* from the distribution $\mathcal{N}(\mathbf{x}_i \beta, \Omega)$
  - Construct the simulated probabilities

$$\tilde{p}_j = \sum_{i=1}^{N} \sum_{r=1}^{R} \mathbf{1}[y^*_{irj} = \max(y^*_{ir0}, y^*_{ir1}, ..., y^*_{irJ})]$$

  - Iterate over β to approximate the true probabilities

Estimation of Multinomial Probit

- **Method of Simulated Moments**
  - Use the first order conditions:

$$\frac{\partial \ln L_N}{\partial \beta} = \sum_i \sum_j \frac{\partial \ln P[y_{ij} = 1 | \mathbf{x}_i]}{\partial \beta}$$

  - Use simulated probabilities and derivatives to solve the first order conditions with respect to β