Multinomial Response Models

(GB: Chapter 15.9-15.10)
R_mlogit_package.pdf

Isabel Casas

Syddansk Universitet

Multinomial data

- Introduction
- Multinomial logit model
- Conditional logit model
- Nested logit model

Introduction

- We consider situations with more than two outcomes.
- If the ordering of the alternatives **is not** meaningful (15.9)
  - Multinomial logit model (MNL)
  - Conditional logit model (CLM)
  - Conditional probit model
  - Nested logit model
- If the ordering of the alternatives **is** meaningful (15.10)
  - Ordered probit model (OPM)
  - Ordered logit model (OLM)

Examples of multinomial response

- Choice between transportation mode: Car, bus or train.
- Choice of political party: V, S, K, R, SF, etc.
- Choice of occupation: wage work; school; self-employment; or out of labour force
- Choice of TV programme: Deadline, Champions League or Paradise Hotel

More than two alternatives $\Rightarrow$ classical Logit and Probit not appropriate.

SYDDANSK UNIVERSITET

Data of multinomial response

The structure of our data is determined by three indexes:

- the alternative
- the choice situation
- the individual

The files can be encountered in two shapes:

- "long" shape: one row for each alternative, as many rows as alternatives for each individual.
- "wide" shape: one row for each choice situation

Data of multinomial response

```
> library(AER)
> data(TravelMode)
> head(TravelMode)

  individual  mode choice wait vcost travel gcost income size
1          1   air     no   69    59    100    70     35    1
2          1 train     no   34    31    372    71     35    1
3          1   bus     no   35    25    417    70     35    1
4          1   car    yes    0    10    180    30     35    1
5          2   air     no   64    58     68    68     30    2
6          2 train     no   44    31    354    84     30    2
```

- Four travel modes (air, train, bus and car),
- most variables are alternative specific (wait, vcost, travel, gcost)
- individual specific variables (income, size).
- This data is in "long" shape.

Data of multinomial response

```
> library(mlogit)
> data(Fishing)
> head(Fishing)

     mode price.beach price.pier price.boat price.charter catch.beach
1 charter     157.930    157.930    157.930       182.930      0.0678
2 charter      15.114     15.114     10.534        34.534      0.1049
3    boat     161.874    161.874     24.334        59.334      0.5333
4    pier      15.134     15.134     55.930        84.930      0.0678
5    boat     106.930    106.930     41.514        71.014      0.0678
6 charter     192.474    192.474     28.934        63.934      0.5333
  catch.pier catch.boat catch.charter   income
1     0.0503     0.2601        0.5391 7083.332
2     0.0451     0.1574        0.4671 1250.000
3     0.4522     0.2413        1.0266 3750.000
4     0.0789     0.1643        0.5391 2083.333
5     0.0503     0.1082        0.3240 4583.332
6     0.4522     0.1665        0.3975 4583.332
```

- Four fishing modes(beach, pier, boat, charter),
- two alternative specific variables (price and catch) and
- one individual specific variable (income).
- This data is in "wide" shape.

Data of multinomial response

```
> library(mlogit)
> data(Train)
> head(Train)

  id choiceid  choice price1 time1 change1 comfort1 price2 time2 change2
1  1        1 choice1   2400   150       0        1   4000   150       0
2  1        2 choice1   2400   150       0        1   3200   130       0
3  1        3 choice1   2400   115       0        1   4000   115       0
4  1        4 choice2   4000   130       0        1   3200   150       0
5  1        5 choice2   2400   150       0        1   3200   150       0
6  1        6 choice2   4000   115       0        0   2400   130       0
  comfort2
1        1
2        1
3        0
4        0
5        0
6        0
```

There is a mix of individual and alternative specific variables:

- modes?
- alternative specific variables?
- individual specific variables?
- wide or long shape?

Multinomial response models

$$P(y_i = j|\mathbf{x}_{ij}, \mathbf{z}_i, w_{ij} = G(\alpha_j + \beta x_{ij} + \gamma_j \mathbf{z}_i + \delta_j w_{ij})$$
$$i = 1, \ldots, n, \quad j = 0, 1, \ldots, J$$

- i is the individual, j is the alternative
- x_{ij} alternative specific variables with generic coefficient β,
- $\mathbf{z}_i$ individual specific variable with an alternative specific coefficient γ_j
- w_{ij} alternative specific variables with an alternative specific coefficient δ_j

Multinomial response models

What is the difference of the probability of two alternatives j, k:

$$\frac{P(y_i = j|\mathbf{x}_i, \mathbf{z}_i, w_i)}{P(y_i = k|\mathbf{x}_i, \mathbf{z}_i, w_i)} = f\,((\alpha_j - \alpha_k) + \beta(x_{ij} - x_{ik}) + (\gamma_j - \gamma_k)\mathbf{z}_i + (\delta_j w_{ij} - \delta_k w_{ik})\,)$$

- Individual specific variables should have alternative specific coefficients because
- If $\alpha_j = \alpha_k$ and $\gamma_j = \gamma_k$ then they would disappear in the differentiation $\Rightarrow$ the alternative would not be different for different individuals.
- Note that only the coefficient differences will be identified.
- So if the alternatives are 1,2,3, then we assume $\alpha_1 = 0$, $\gamma_1 = 0$

SYDDANSK UNIVERSITET

Multinomial response models

What is the difference of the probability of two alternatives j, k:

$$\frac{P(y_i = j|\mathbf{x}_i, \mathbf{z}_i, w_i)}{P(y_i = k|\mathbf{x}_i, \mathbf{z}_i, w_i)} = f\left((\alpha_j - \alpha_k) + \beta(x_{ij} - x_{ik}) + (\gamma_j - \gamma_k)\mathbf{z}_i + (\delta_j w_{ij} - \delta_k w_{ik})\right)$$

- Alternative specific variables may or may not have alternative specific coefficients
- Travel time is an alternative specific variable and 10 minutes in a train might not have the same impact than 10 minutes in a car. Then, we would use an alternative specific coefficient δ_j.
- Monetary time travel is also alternative specific variable, however one euro spent in time travel is the same independently of the type of transportation, so then a constant β makes sense.

Multinomial response models

- A model with only individual specific variables is called a *multinomial logit model*:

$$P(y_i = j|\mathbf{z}_i) = G(\alpha_j + \gamma_j \mathbf{z}_i), \quad i = 1, \ldots, n, \quad j = 0, 1, \ldots, J$$

- A model with only alternative specific variables is called a *conditional logit model*:

$$P(y_i = j|\mathbf{x}_i, w_i) = G(\alpha_j + \beta x_{ij} + \delta_j w_{ij})$$

  Note that this term is misleading.
- One with both kind of variables is called a *mixed logit model*:

$$P(y_i = j|\mathbf{x}_i, \mathbf{z}_i, w_i) = G(\alpha_j + \beta x_{ij} + \gamma_j \mathbf{z}_i + \delta_j w_{ij})$$

Multinomial Logit Model (MNL)

- Model set–up
- Interpretation of γ_j
  - Marginal effects of $\mathbf{z}_k$
  - Log–odd ratio
- Reporting results
- R command
- Example

Model for individual–specific data

- y takes a value in $\{0, 1, 2, \ldots, J\}$
- For example $y = 0, 1, 2, 3$ indicating occupational choice: private company; self–employment; academic; out of labour force.
- These values are nominal
- $\mathbf{Z}$ is $n \times (k+1)$ matrix of individual characteristics
- For example, $\mathbf{Z}$ contains: 1, education, age, race and marital status
- A sample of n random draws: $(y_i, \mathbf{z}_i), i = 1, \ldots, n$ from the population
- n_j is the number of observations with response $y_i = j$
- The total number of observations $n = \sum_{j=0}^{n} n_j$

SYDDANSK UNIVERSITET

Model for individual–specific data

- We are interested in the response probability:

$$p_j = P(y = j|\mathbf{Z}), \quad j = 0, 1, \ldots, J$$

- We are interested in how the changes in the elements of $\mathbf{Z}$ affect the probabilities of the response
- Note that we need only know J of these outcomes, Why?

Model set–up

Let us assume that there are $J+1$ possible outcomes.

$$P(y=j|\mathbf{Z}) = \frac{\exp(\mathbf{Z}\gamma_j)}{1+\sum_{h=1}^{J}\exp(\mathbf{Z}\gamma_h)}, \quad j=1,\ldots,J$$

where $\gamma_j = (\gamma_{j0}, \gamma_{j1}, \ldots, \gamma_{jk})$, for k number of variables in $\mathbf{Z}$.

Because $P(y=0|\mathbf{Z}) + P(y=1|\mathbf{Z}) + \ldots + P(y=J|\mathbf{Z}) = 1$,

$$P(y=0|\mathbf{Z}) = \frac{1}{1+\sum_{h=1}^{J}\exp(\mathbf{Z}\gamma_h)}$$

Basically $\gamma_{00} = \gamma_{01} = \ldots = \gamma_{0k} = 0$

If $J=1 \Rightarrow$ we have the Binary logit model.

Partial effects of continuous $\mathbf{z}_k$

The partial (marginal) effect for a continuous variable $\mathbf{z}_k$:

$$\frac{\partial P(y=j|\mathbf{Z})}{\partial \mathbf{z}_k} = \frac{\exp(\mathbf{Z}\gamma_j)\gamma_{jk}\left[1+\sum_{h=1}^{J}\exp(\mathbf{Z}\gamma_h)\right] - \exp(\mathbf{Z}\gamma_j)\sum_{h=1}^{J}\gamma_{hk}\exp(\mathbf{Z}\gamma_h)}{\left[1+\sum_{h=1}^{J}\exp(\mathbf{Z}\gamma_h)\right]^2}$$

$$= \frac{\exp(\mathbf{Z}\gamma_j)\gamma_{jk}}{\left[1+\sum_{h=1}^{J}\exp(\mathbf{Z}\gamma_h)\right]} - \frac{\exp(\mathbf{Z}\gamma_j)\sum_{h=1}^{J}\gamma_{hk}\exp(\mathbf{Z}\gamma_h)}{\left[1+\sum_{h=1}^{J}\exp(\mathbf{Z}\gamma_h)\right]^2}$$

$$= P(y=j|\mathbf{Z})\left[\gamma_{jk} - \frac{\sum_{h=1}^{J}\gamma_{hk}\exp(\mathbf{Z}\gamma_h)}{\left[1+\sum_{h=1}^{J}\exp(\mathbf{Z}\gamma_h)\right]}\right], \quad j=1,2,\ldots,J$$

Exercise: How is the partial effect of a continuous variable $\mathbf{z}_k$ over $P(y=0|\mathbf{Z})$? (5 minutes)

Partial effects of continuous $\mathbf{z}_k$

$$\frac{\partial P(y=j|\mathbf{Z})}{\partial \mathbf{z}_k} = P(y=j|\mathbf{Z}) \left[\gamma_{jk} - \frac{\sum_{h=1}^{J} \gamma_{hk} \exp(\mathbf{Z}\gamma_h)}{\left[1 + \sum_{h=1}^{J} \exp(\mathbf{Z}\gamma_h)\right]} \right]$$

- $\mathbf{z}_k$ is one of the variables in $\mathbf{Z}$
- γ_{jk} is the coefficient of $\mathbf{z}_k$ for alternative j
- The sign of the partial effect $\neq$ sign of γ_{jk}

Interpretation of γ_j: odds ratio

Odds ratio or relative risk is defined as:

$$\frac{P(y=j|\mathbf{Z})}{P(y=0|\mathbf{Z})} = \exp(\mathbf{Z}\gamma_j) = \exp(\gamma_{j0} + \gamma_{j1}\mathbf{z}_1 + \ldots + \gamma_{jk}\mathbf{z}_k)$$

- E.g. $\exp(\gamma_{j1})$ expresses how much likely (or unlikely) is event $y=j$ to happen than event $y=0$ when $\mathbf{z}_1$ changes
- If $\exp(\gamma_{j1})$=1 then the two events are equally likely when $\mathbf{z}_1$ changes
- If $\exp(\gamma_{j1}) > 1$ then event $y=j$ is more likely to occur than event $y=0$
- If $\exp(\gamma_{j1}) < 1$ then event $y=j$ is less likely to occur than event $y=0$

Syddansk Universitet

Interpretation of γ_j: log-odds ratio

A simpler interpretation of γ_j is given by the relative probability (the odds ratio) between two alternatives:

$$\log\left(\frac{P(y=j|\mathbf{Z})}{P(y=0|\mathbf{Z})}\right) = \mathbf{Z}\gamma_j = \gamma_{j0} + \gamma_{j1}\mathbf{z}_1 + \ldots + \gamma_{jk}\mathbf{z}_k$$

- E.g. γ_{j1} expresses how much likely (or unlikely) is event $y=j$ to happen than event $y=0$ when $\mathbf{z}_1$ changes
- If γ_{j1}=0 then the two events are equally likely when $\mathbf{z}_1$ changes
- If $\gamma_{j1} > 0$ then event $y=j$ is more likely to occur than event $y=0$
- If $\gamma_{j1} < 0$ then event $y=j$ is less likely to occur than event $y=0$

Interpretation of γ_j

Also,

$$\log\left(\frac{P(y=j|\mathbf{Z})}{P(y=p|\mathbf{Z})}\right)=\mathbf{Z}(\gamma_j-\gamma_p)$$

- If $(\gamma_{j1}-\gamma_{p1})=0$ expresses that event $y=j$ and $y=p$ are equally likely when variable $\mathbf{z}_1$ changes
- $(\gamma_{j1}-\gamma_{p1})>0$ expresses that event $y=j$ is more likely than event $y=p$ when variable $\mathbf{z}_1$ changes
- $(\gamma_{j1}-\gamma_{p1})<0$ ...

Syddansk Universitet

Example: Interpretation of γ_j

Example: 3 outcomes:

- $y = 0$ (in school)
- $y = 1$ (not in the school and not working, at home)
- $y = 2$ (working)
- Explanatory variables: $\mathbf{z}_1$=experience, $\mathbf{z}_2$=education

How should we interpret:

- $\gamma_{11} < 0$ Another year of experience reduces the log–odds between at home and enrolled in school. The probability of event "at home" decreases in comparison to the probability of event " in school" as the number of years of experience increase
- $\gamma_{12} > 0$ Another year of education increases the log–odds between at home and enrolled in school.

Maximum Likelihood Estimation of MNL

Log-likelihood for observation i is:

$$\ell_i(\gamma) = \sum_{j=0}^{J} \mathbf{1}[y_i = j] \log p_j(\mathbf{z}_i, \gamma)$$

where

- E.g. $p_2(\mathbf{z}_i, \gamma) = p_2(\mathbf{z}_i, \gamma_0, \gamma_1, \ldots, \gamma_J)$ is the prob. that $y_i = 2$ given $\mathbf{z}_i$
- $\mathbf{1}[y_i = 2] = 1$ if $y_i = 2$ and zero other ways,
- This indicator function "picks" the right probability for observation i for the likelihood function
- Probability of alternative j for observation i

Maximum Likelihood Estimation of MNL

The maximum likelihood estimator maximises the likelihood function:

$$\ell(\gamma) = \sum_{i=1}^{n} \ell_i(\gamma) = \sum_{i=1}^{n} \sum_{j=0}^{J} \mathbf{1}[y_i = j] \log p_j(\mathbf{z}_i, \gamma)$$

w.r.t γ

- If the ML estimation conditions are satisfied then these estimates of MNL are
  - Consistent, asymptotically normal and asymptotically efficient
- Same tests: Wald, LR and LM
- The asymptotic variance can be estimated with any of the three estimators we saw before
- The same consequences of misspecification hold for this estimator.

Example 15.4 Wooldridge

- status: school=1,home=2,work=3
- $educ$, $exper$, $expersq$, $black$ (individual specific variables)
- Data from 1987

	status	educ	exper	expersq	black
1	2	10	0	0	1
2	2	10	0	0	1
3	2	10	0	0	1
4	1	10	0	0	1
5	2	11	0	0	1
6	2	11	0	0	1
7	2	11	0	0	1
8	2	12	0	0	1
9	3	12	0	0	1
10	3	12	1	1	1

R commands

- Install packages mlogit
- Load package mlogit
- Read the data file
- Make a data.frame with the variables of interest: status, educ, exper, expersq, black, id
- Convert this file in a format that mlogit can understand (function mlogit.data)
- Understand what formula you need to use
- Use mlogit to estimate the model

R commands

- Read file

```
> keane<-read.table("keane.raw", na.string=".")
```

- We want data from 1987 and we want to get rid of the missing values (note .RAW file does not have headers)
- From the .des file, we see that variable number 6 is educ, variable number 19 is y87, variable number 23 is exper, variable 24 is expersq and variable 25 is status.

```
> index<-which(keane[,19]!=1 | is.na(keane[,25]) | is.na(keane[,6])|
is.na(keane[,23]) | is.na(keane[,24]))
> keane<-keane[-index,]
```

R commands

- Create a data frame with our variables

```
> keane2<-data.frame(id= keane[,1],status=keane[,25],
educ=keane[,6], exper=keane[,23], expersq=keane[,24],
black=keane[,11])
```

- Are our variables alternative specific?

R commands

- Install mlogit package
- Load mlogit package

```
> library(mlogit)
```

- Write the data in a format the mlogit understand
- The variable exper, educ, expersq, black are individual specific
- The response variable is "status" so it goes into the parameter "choice"
- shape: wide (one row for each alternative)

```
> keane3<-mlogit.data(keane2, choice = "status", shape="wide", id="id")
```

R commands

How do we write the formula for mlogit to understand we are using a multinomial logit model, a conditional logit model or a mixed logit model?

The right hand side of the formula contains three parts:

1. alternative specific variables with generic coefficients (βx_{ij})
2. individual specific variables with alternative specific coefficients $(\gamma_j \mathbf{z}_i, \alpha_j)$. By default $\alpha_1 = \gamma_1 = 0$.
3. alternative specific variableswith alternative specific coefficients $(\delta_j w_{ij})$

$$choice \sim x_{ij} | \mathbf{z}_i | w_{ij}$$

by default α_j is included unless we write 0, -1 in the middle part of the formula.

R commands

Examples of formula for the MNL $(\alpha_j + \gamma_j \mathbf{Z})$

$$
\begin{aligned}
&choice \sim 0|income|0 \\
&choice \sim 0|-1 + income|0 \\
&choice \sim 0|income) \\
&choice \sim 0|-1 + income)
\end{aligned}
$$

R commands

- Estimate the parameters of a MNL using *mlogit*.
- status: school=1,home=2,work=3
- individual specific variables: exper, educ, expersq, black
- reflevel="1", the one who will have parameters $\alpha_1 = 0 = \gamma_1$

```
> keane.model<-mlogit(status~0| educ+exper+expersq+black,
            data=keane3, reflevel="1")
> summary(keane.model)
```

R commands

```
Coefficients :
                Estimate Std. Error t-value  Pr(>|t|)
2:(intercept) 10.277874   1.133336  9.0687 < 2.2e-16 ***
3:(intercept)  5.543797   1.086409  5.1029 3.346e-07 ***
2:educ        -0.673631   0.069900 -9.6371 < 2.2e-16 ***
3:educ        -0.314657   0.065110 -4.8327 1.347e-06 ***
2:exper       -0.106215   0.173282 -0.6130 0.5399029
3:exper        0.848737   0.156986  5.4065 6.428e-08 ***
2:expersq     -0.012515   0.025229 -0.4961 0.6198523
3:expersq     -0.077300   0.022922 -3.3724 0.0007453 ***
2:black        0.813017   0.302723  2.6857 0.0072383 **
3:black        0.311361   0.281534  1.1059 0.2687500
---

Log-Likelihood: -907.86
McFadden R^2:  0.24327
Likelihood ratio test : chisq = 583.72 (p.value = < 2.22e-16)
```

SYDDANSK UNIVERSITET

Reporting results of MNL

- status: school=1,home=2,work=3
- Log–odds ratios (relative to alternative status=1, it is just the parameters)

$$log\left(\frac{P(status=2)}{P(status=1)}\right) = \gamma_{20} + \gamma_{21}\ educ + \gamma_{22}\ exper + \gamma_{23}\ exper^2 + \gamma_{24}\ black$$
$$log\left(\frac{P(status=3)}{P(status=1)}\right) = \gamma_{30} + \gamma_{31}\ educ + \gamma_{32}\ exper + \gamma_{33}\ exper^2 + \gamma_{34}\ black$$

- E.g. for one unit change in the variable $educ$:
  - the log of the ratio of the two probabilities $\gamma_{21} = -0.67$,
  - The log of the ratio of the two probabilities $\gamma_{31} = -0.31$.
- Therefore the probability of being at home or at work decrease as the years of education increases.

Reporting results of MNL

- The ratio of the probability of choosing one outcome alternative over the probability of choosing the reference alternative is often referred as relative risk.

```
> exp(coef(keane.model))
        alt2         alt3    alt2:educ    alt3:educ   alt2:exper
2.908198e+04 2.556469e+02 5.098538e-01 7.300390e-01 8.992314e-01
 alt3:exper alt2:expersq alt3:expersq   alt2:black   alt3:black
2.336693e+00 9.875628e-01 9.256118e-01 2.254699e+00 1.365282e+00
```

- For one unit change in the variable $educ$, we expect the relative risk of choosing $status = 2$ over $status = 1$ to decrease 0.5.
- So, the relative risk is lower for more educated people.
- For a dummy variable such as $black$: the ratio of the relative risks of choosing $status = 2$ over $status = 1$ for black is 2.25. So there is a higher probability to stay at home than to be at school.

Reporting results of MNL

- Percent correctly predicted observations
  - In total
  - Separately for each of the $J+1$ outcomes.
- Differences in fitted probabilities between two typical (or relevant) observations of $\mathbf{Z}$
  - E.g. the average male and the average female person