ECON 4551
Econometrics II
Memorial University of Newfoundland

Qualitative and Limited Dependent Variable Models

Chapter 16

Adapted from Vera Tabakova's notes

Chapter 16: Qualitative and Limited Dependent Variable Models

16.3 Multinomial Logit

Examples of multinomial choice (***polytomous***) situations:

1. Choice of a laundry detergent: Tide, Cheer, Arm & Hammer, Wisk, etc.
2. Choice of a major: economics, marketing, management, finance or accounting.
3. Choices after graduating from high school: not going to college, going to a private 4-year college, a public 4 year-college, or a 2-year college.

16.3 Multinomial Logit

The explanatory variable x_i is **individual specific**, but does not change across alternatives. Example age of the individual.

The dependent variable is ***nominal***

16.3 Multinomial Logit

Examples of multinomial choice situations:

1. It is key that there are more than 2 choices

2. It is key that there is no meaningful ordering to them. Otherwise we would want to use that information (with an ordered probit or ordered logit)

   In GRETL, you must specify MNL, or the program will assume meaningful ordering!

16.3 Multinomial Logit

In essence this model is like a set of simultaneous individual binomial logistic regressions

With appropriate weighting, since the different comparisons between different pairs of categories would generally involve different numbers of observations

16.3.1 Multinomial Logit Choice Probabilities

$$p_{ij} = P\left[\text{individual } i \text{ chooses alternative } j\right]$$

Why is this "one"

$$p_{i1} = \frac{1}{1+\exp\left(\beta_{12}+\beta_{22}x_i\right)+\exp\left(\beta_{13}+\beta_{23}x_i\right)}, \; j=1 \qquad (16.19a)$$

$$p_{i2} = \frac{\exp\left(\beta_{12}+\beta_{22}x_i\right)}{1+\exp\left(\beta_{12}+\beta_{22}x_i\right)+\exp\left(\beta_{13}+\beta_{23}x_i\right)}, \; j=2 \qquad (16.19b)$$

$$p_{i3} = \frac{\exp\left(\beta_{13}+\beta_{23}x_i\right)}{1+\exp\left(\beta_{12}+\beta_{22}x_i\right)+\exp\left(\beta_{13}+\beta_{23}x_i\right)}, \; j=3 \qquad (16.19c)$$

16.3.2 Maximum Likelihood Estimation

$$P\left[y_{11}=1, y_{22}=1, y_{33}=1\right] = p_{11}\times p_{22}\times p_{33}$$

$$= \frac{1}{1+\exp\left(\beta_{12}+\beta_{22}x_1\right)+\exp\left(\beta_{13}+\beta_{23}x_1\right)}\times$$

We solve using Maximum Likelihood

$$\frac{\exp\left(\beta_{12}+\beta_{22}x_2\right)}{1+\exp\left(\beta_{12}+\beta_{22}x_2\right)+\exp\left(\beta_{13}+\beta_{23}x_2\right)}\times$$

$$\frac{\exp\left(\beta_{13}+\beta_{23}x_3\right)}{1+\exp\left(\beta_{12}+\beta_{22}x_3\right)+\exp\left(\beta_{13}+\beta_{23}x_3\right)}$$

$$= L\left(\beta_{12},\beta_{22},\beta_{13},\beta_{23}\right)$$

16.3.3 Post-Estimation Analysis

Again, marginal effects are complicated: there are several types of reporting to consider

$$\tilde{p}_{01} = \frac{1}{1+\exp\left(\tilde{\beta}_{12}+\tilde{\beta}_{22}x_0\right)+\exp\left(\tilde{\beta}_{13}+\tilde{\beta}_{23}x_0\right)}$$

$$\left.\frac{\Delta p_{im}}{\Delta x_i}\right|_{\text{all else constant}} = \frac{\partial p_{im}}{\partial x_i} = p_{im}\left[\beta_{2m} - \sum_{j=1}^{3}\beta_{2j}p_{ij}\right] \qquad (16.20)$$

For example reporting the difference in predicted probabilities for two values of a variable

$$\Delta p_1 = \tilde{p}_{b1} - \tilde{p}_{a1}$$

$$= \frac{1}{1+\exp\left(\tilde{\beta}_{12}+\tilde{\beta}_{22}x_b\right)+\exp\left(\tilde{\beta}_{13}+\tilde{\beta}_{23}x_b\right)} - \frac{1}{1+\exp\left(\tilde{\beta}_{12}+\tilde{\beta}_{22}x_a\right)+\exp\left(\tilde{\beta}_{13}+\tilde{\beta}_{23}x_a\right)}$$

16.3.3 Post-Estimation Analysis: odds ratios

$$\frac{P(y_i = j)}{P(y_i = 1)} = \frac{p_{ij}}{p_{i1}} = \exp\left(\beta_{1j} + \beta_{2j} x_i\right) \quad j = 2,3 \tag{16.21}$$

$$\frac{\partial\left(p_{ij} / p_{i1}\right)}{\partial x_i} = \beta_{2j} \exp\left(\beta_{1j} + \beta_{2j} x_i\right) \quad j = 2,3 \tag{16.22}$$

An interesting feature of the odds ratio (16.21) is that the odds of choosing alternative *j* rather than alternative 1 does not depend on how many alternatives there are in total. There is the implicit assumption in logit models that the odds between any pair of alternatives is **independent of irrelevant alternatives (IIA)**.

IIA assumption

- There is the implicit assumption in logit models that the odds between any pair of alternatives is **independent of irrelevant alternatives (IIA)**

One way to state the assumption

- *If choice A is preferred to choice B out of the choice set {A,B}, then introducing a third alternative X, thus expanding that choice set to {A,B,X}, must not make B preferable to A.*

- which kind of makes sense ☺

IIA assumption

- There is the implicit assumption in logit models that the odds between any pair of alternatives is **independent of irrelevant alternatives (IIA)**

In the case of the multinomial logit model, the IIA implies that adding another alternative or changing the characteristics of a third alternative must not affect the relative odds between the two alternatives considered.

This is not realistic for many real life applications involving similar (***substitute***) alternatives.

IIA assumption

This is not realistic for many real life applications with similar (substitute) alternatives

Examples:

- Beethoven/Debussy versus another of Beethoven's Symphonies (Debreu 1960; Tversky 1972)
- Bicycle/Pony (Luce and Suppes 1965)
- Red Bus/Blue Bus (McFadden 1974).
- Black slacks, jeans, shorts versus blue slacks (Hoffman, 2004)
- Etc.

IIA assumption

Red Bus/Blue Bus (McFadden 1974).

- Imagine commuters first face a decision between two modes of transportation: ***car*** and ***red bus***
- Suppose that a consumer chooses between these two options with equal probability, 0.5, so that the odds ratio equals 1.
- Now add a third mode, ***blue bus***. Assuming bus commuters do not care about the color of the bus (they are perfect substitutes), consumers are expected to choose between bus and car still with equal probability, so the probability of car is still 0.5, while the probabilities of each of the two bus types should go down to 0.25
- However, this violates IIA: for the odds ratio between car and red bus to be preserved, the new probabilities must be: ***car*** 0.33; ***red bus*** 0.33; ***blue bus*** 0.33
- Te IIA axiom does not mix well with perfect substitutes ☹

IIA assumption

We can test this assumption with a Hausman-McFadden test which compares a logistic model with all the choices with one with restricted choices (***mlogtest, hausman base*** in STATA, but check option detail too: ***mlogtest, hausman detail***)

However, see [Cheng and Long (2007)]

Another test is Small and Hsiao's (1985)
STATA's command is ***mlogtest, smhsiao*** **(**careful: the sample is randomly split every time, so you must set the *seed* if you want to replicate your results**)**

See Long and Freese's book for details and worked examples

IIA assumption

use nels_small, clear

```
. mlogit psechoice grades  faminc parcoll, baseoutcome(1) nolog

Multinomial logistic regression                   Number of obs   =       1000
                                                  LR chi2(6)      =     342.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -847.54576                       Pseudo R2       =     0.1680

------------------------------------------------------------------------------
   psechoice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |  (base outcome)
-------------+----------------------------------------------------------------
2            |
      grades |  -.2891448   .0530752    -5.45   0.000    -.3931703   -.1851192
      faminc |   .0080757    .004009     2.01   0.044     .0002182    .0159332
     parcoll |   .5370023   .2892469     1.86   0.063    -.0299112    1.103916
       _cons |   1.942856   .4561356     4.26   0.000     1.048847    2.836866
-------------+----------------------------------------------------------------
3            |
      grades |  -.6558358   .0540845   -12.13   0.000    -.7618394   -.5498321
      faminc |   .0132383   .0038992     3.40   0.001      .005596    .0208807
     parcoll |   1.067561    .274181     3.89   0.000     .5301758    1.604946
       _cons |    4.57382   .4392376    10.41   0.000      3.71293     5.43471
------------------------------------------------------------------------------
```

average grade on 13 point scale with 1 = highest

```
. mlogtest, hausman

**** Hausman tests of IIA assumption (N=1000)

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

 Omitted |      chi2   df   P>chi2   evidence
---------+------------------------------------
       2 |     0.206    4    0.995   for Ho
       3 |     0.021    4    1.000   for Ho
----------------------------------------------
```

IIA assumption

```
. mlogit psechoice grades  faminc  , baseoutcome(1) nolog

Multinomial logistic regression                 Number of obs     =       1000
                                                LR chi2(4)        =     323.70
                                                Prob > chi2       =     0.0000
Log likelihood = -856.80718                     Pseudo R2         =     0.1589

------------------------------------------------------------------------------
   psechoice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |  (base outcome)
-------------+----------------------------------------------------------------
2            |
      grades |  -.2962217   .0526424    -5.63   0.000    -.3993989   -.1930446
      faminc |   .0108711   .0038504     2.82   0.005     .0033245    .0184177
       _cons |   1.965071   .4550879     4.32   0.000     1.073115    2.857027
-------------+----------------------------------------------------------------
3            |
      grades |  -.6794793   .0535091   -12.70   0.000    -.7843553   -.5746034
      faminc |   .0188675   .0037282     5.06   0.000     .0115603    .0261747
       _cons |   4.724423   .4362826    10.83   0.000     3.869325    5.579521
------------------------------------------------------------------------------

. mlogtest, smhsiao

**** Small-Hsiao tests of IIA assumption (N=1000)

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

 Omitted |  lnL(full)  lnL(omit)    chi2   df   P>chi2   evidence
---------+---------------------------------------------------------
       2 |   -171.559   -170.581   1.955    3    0.582   for Ho
       3 |   -156.227   -153.342   5.770    3    0.123   for Ho
-------------------------------------------------------------------

.
```

IIA assumption

The randomness…due to different splittings of the sample

```
. mlogtest, smhsiao

**** Small-Hsiao tests of IIA assumption (N=1000)

 Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

 Omitted |   lnL(full)  lnL(omit)    chi2    df   P>chi2    evidence
---------+-----------------------------------------------------------
       2 |    -158.961   -154.880   8.162     3    0.043    against Ho
       3 |    -149.106   -147.165   3.880     3    0.275    for Ho
---------------------------------------------------------------------
```

IIA assumption

- Extensions have arisen to deal with this issue
- The multinomial probit and the mixed logit are alternative models for nominal outcomes that relax IIA, by allowing correlation among the errors (to reflect similarity among options)
- but these models often have issues and assumptions themselves ☹
- IIA can also be relaxed by specifying a hierarchical model, ranking the choice alternatives. The most popular of these is called the McFadden's ***nested logit model,*** which allows correlation among some errors, but not all (e.g. Heiss 2002)
- Generalized extreme value and multinomial probit models possess another property, the ***Invariant Proportion of Substitution*** (Steenburgh 2008), which itself also suggests similarly counterintuitive real-life individual choice behavior
- The multinomial probit has serious computational disadvantages too, since it involves calculating multiple (one less than the number of categories) integrals. With integration by simulation this problem is being ameliorated now…

16.3.4 An Example

```
. tab psechoice
```

no college = 1, 2 = 2-year college, 3 = 4-year college	Freq.	Percent	Cum.
1	222	22.20	22.20
2	251	25.10	47.30
3	527	52.70	100.00
Total	1,000	100.00	

16.3.4 An Example

mlogit psechoice grades, baseoutcome(1)

```
. mlogit psechoice grades, baseoutcome(1)

Iteration 0:   log likelihood = -1018.6575
Iteration 1:   log likelihood = -881.68524
Iteration 2:   log likelihood = -875.36084
Iteration 3:   log likelihood = -875.31309
Iteration 4:   log likelihood = -875.31309

Multinomial logistic regression                   Number of obs   =       1000
                                                  LR chi2(2)      =     286.69
                                                  Prob > chi2     =     0.0000
Log likelihood = -875.31309                       Pseudo R2       =     0.1407

------------------------------------------------------------------------------
   psechoice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |  (base outcome)
-------------+----------------------------------------------------------------
2            |
      grades |  -.3087889   .0522849    -5.91   0.000    -.4112654   -.2063125
       _cons |   2.506421   .4183848     5.99   0.000     1.686402     3.32644
-------------+----------------------------------------------------------------
3            |
      grades |  -.7061967   .0529246   -13.34   0.000     -.809927   -.6024664
       _cons |   5.769876   .4043229    14.27   0.000     4.977417    6.562334
------------------------------------------------------------------------------
```

16.3.4 An Example

. ***tab psechoice, gen(coll)***

So we can run the individual logits by hand…here "3-year college" versus "no college"

```
. logit  coll2  grades if  psechoice<3

Iteration 0:   log likelihood = -326.96905
Iteration 1:   log likelihood = -308.40836
Iteration 2:   log likelihood = -308.37104
Iteration 3:   log likelihood = -308.37104

Logistic regression                               Number of obs   =        473
                                                  LR chi2(1)      =      37.20
                                                  Prob > chi2     =     0.0000
Log likelihood = -308.37104                       Pseudo R2       =     0.0569

------------------------------------------------------------------------------
       coll2 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      grades |  -.3059161    .053113    -5.76   0.000    -.4100156   -.2018165
       _cons |   2.483675   .4241442     5.86   0.000     1.652367    3.314982
------------------------------------------------------------------------------
```

16.3.4 An Example

. ***tab psechoice, gen(coll)***

So we can run the individual logits by hand…here “4 year college” versus “no college”

```
. logit  coll3  grades if  psechoice!=2

Iteration 0:   log likelihood = -455.22643
Iteration 1:   log likelihood = -337.82899
Iteration 2:   log likelihood = -328.85866
Iteration 3:   log likelihood = -328.76478
Iteration 4:   log likelihood = -328.76471
Iteration 5:   log likelihood = -328.76471

Logistic regression                               Number of obs   =        749
                                                  LR chi2(1)      =     252.92
                                                  Prob > chi2     =     0.0000
Log likelihood = -328.76471                       Pseudo R2       =     0.2778

------------------------------------------------------------------------------
       coll3 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      grades |  -.7151864    .0576598   -12.40   0.000    -.8281976   -.6021752
       _cons |   5.832757     .436065    13.38   0.000     4.978085    6.687428
------------------------------------------------------------------------------
```

Coefficients should look familiar…

But check sample sizes!

16.3.4 An Example

Table 16.2 Maximum Likelihood Estimates of PSE Choice

Parameters	Estimates	Standard errors	t-Statistics
β_{12}	2.5064	0.4183	5.99
β_{22}	−0.3088	0.0523	−5.91
β_{13}	5.7699	0.4043	14.27
β_{23}	−0.7062	0.0529	−13.34

16.3.4 An Example

Table 16.3 **Effects of Grades on Probability of PSE Choice**

PSE choice	*GRADES*	$\hat{p}$	Marginal effect
No college	6.64	0.181	0.084
	2.635	0.018	0.012
2-Year college	6.64	0.286	0.045
	2.635	0.097	0.033
4-Year college	6.64	0.533	−0.128
	2.635	0.886	−0.045

16.3.4 An Example

```
* compute predictions and summarize
predict ProbNo ProbCC ProbColl
summarize ProbNo ProbCC ProbColl
```

```
. predict ProbNo ProbCC ProbColl
(option pr assumed; predicted probabilities)

. summarize ProbNo ProbCC ProbColl
```

This must always
Happen, so do not
Use sample values
To assess predictive accuracy!

Variable	Obs	Mean	Std. Dev.	Min	Max
ProbNo	1000	.222	0	.222	.222
ProbCC	1000	.251	0	.251	.251
ProbColl	1000	.527	0	.527	.527

16.3.4 An Example

Compute marginal effects, say for outcome 1 (no college)

If not specified, calculation is done at means

```
. mfx, predict(outcome(1))

Marginal effects after mlogit
      y  = Pr(psechoice==1) (predict, outcome(1))
         =  .17193474
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
grades	.0813688	.00595	13.68	0.000	.069707	.09303	6.53039

```
. sum grades
```

Variable	Obs	Mean	Std. Dev.	Min	Max
grades	1000	6.53039	2.265855	1.74	12.33

16.3.4 An Example

Compute marginal effects, say for outcome 1 (no college)

If specified, calculation is done at chosen level

```
. mfx, predict(outcome(1)) at ( grades=5)

Marginal effects after mlogit
      y  = Pr(psechoice==1) (predict, outcome(1))
         =  .07691655
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
  grades |   .0439846       .00357   12.31   0.000   .036984  .050985        5
------------------------------------------------------------------------------
```

Example

Another annotated example

- http://www.ats.ucla.edu/stat/Stata/output/stata_mlogit_output.htm

- This example showcases also the use of the option rrr which yields the interpretation of the multinomial logistic regression in terms of relative risk ratios

- In general, the relative risk is a ratio of the probability of an event in the exposed group versus a non-exposed group. Used often in epidemiology

Example

In STATA

- ***mlogit***
- Note that you should specify the base category or STATA will choose the most frequent one
- It is interesting to experiment with changing the base category
- Or use ***listcoef*** to get more results automatically

Example

In STATA

Careful with perfect prediction, which in this model will not be flagged!!!

You can see that the Z values are zero for some variables and the p-values will be 1, but STATA will not send a warning message now!

Similar for ologit and oprobit later…

Combining Categories

Consider testing whether two categories could be combined

If none of the independent variables really explain the odds of choosing choice A versus B, you should merge them

In STATA

mlogtest, combine (Wald test)

Or

mlogtest, lrcomb (LR test)

Our example…

mlogit psechoice grades faminc , baseoutcome(3)

```
. mlogtest, combine

**** Wald tests for combining alternatives (N=1000)

 Ho: All coefficients except intercepts associated with a given pair
     of alternatives are 0 (i.e., alternatives can be combined).

Alternatives tested|      chi2    df    P>chi2
-------------------+----------------------------
       1-        2 |    41.225     2     0.000
       1-        3 |   187.029     2     0.000
       2-        3 |    97.658     2     0.000
------------------------------------------------
```

Where does this come from?

Our example…

mlogit psechoice grades faminc , baseoutcome(3)

```
. test[1]

 ( 1)  [1]grades = 0
 ( 2)  [1]faminc = 0

           chi2(  2) =  187.03
         Prob > chi2 =    0.0000
```

We test whether all the
Coefficients are null
When comparing
category 1 to the base,
Which is 3 here

Our example…

mlogit psechoice grades faminc , baseoutcome(3)

```
. mlogtest, lrcomb

**** LR tests for combining alternatives (N=1000)

 Ho: All coefficients except intercepts associated with a given pair
     of alternatives are 0 (i.e., alternatives can be collapsed).

Alternatives tested|      chi2    df   P>chi2
-------------------+----------------------------
       1-        2 |    46.360     2    0.000
       1-        3 |   294.004     2    0.000
       2-        3 |   118.271     2    0.000
```

These tests are based on comparing unrestricted versus constrained Regressions, where only the intercept is nonzero for the relevant category

Our example...

These tests are based on comparing unrestricted versus constrained Regressions, where only the intercept is nonzero for the relevant category:

mlogit psechoice grades faminc , baseoutcome(3) nolog
est store unrestricted
constraint define 27 [1]
mlogit psechoice grades faminc , baseoutcome(3) constraint(27) nolog
est store restricted
lrtest restricted unrestricted

Yields:

```
Likelihood-ratio test                                 LR chi2(2)  =    294.00
(Assumption: restricted nested in unrestricted)       Prob > chi2 =    0.0000
```

Multinomial Logit versus Probit

Computational issues make the Multinomial Probit very rare

LIMDEP seemed to be one of the few software packages that used to include a canned routine for it

STATA has now ***asmprobit***

Advantage: it does not need IIA ☺

Logit as special case of Multinomial Logit

```
. tab  hscath

     = 1 if |
   catholic |
high school |
   graduate |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        981       98.10       98.10
          1 |         19        1.90      100.00
------------+-----------------------------------
      Total |      1,000      100.00
```

Logit as special case of Multinomial Logit

```
. mlogit  hscath grades, baseoutcome(1) nolog
```

Multinomial logistic regression

Number of obs	=	1000
LR chi2(1)	=	0.21
Prob > chi2	=	0.6445
Pseudo R2	=	0.0011

Log likelihood = -94.014874

hscath	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
0						
grades	.0471052	.1020326	0.46	0.644	-.1528749	.2470853
_cons	3.642004	.6830122	5.33	0.000	2.303325	4.980684
1	(base outcome)					

Why are the coefficient signs reversed?

```
. logit  hscath grades, nolog
```

Logistic regression

Number of obs	=	1000
LR chi2(1)	=	0.21
Prob > chi2	=	0.6445
Pseudo R2	=	0.0011

Log likelihood = -94.014874

hscath	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
grades	-.0471052	.1020326	-0.46	0.644	-.2470853	.1528749
_cons	-3.642004	.6830122	-5.33	0.000	-4.980684	-2.303325

16.4 Conditional Logit

Example: choice between three types ($J = 3$) of soft drinks, say Pepsi, 7-Up and Coke Classic.

Let y_{i1}, y_{i2} and y_{i3} be dummy variables that indicate the choice made by individual i. The price facing individual i for brand j is $PRICE_{ij}$.

Variables like price are to be **individual and alternative specific**, because they vary from individual to individual and are different for each choice the consumer might make

16.4 Conditional Logit

Variables like price are to be **individual <u>and</u> alternative specific**, because they vary from individual to individual and are different for each choice the consumer might make

Another example: of mode of transportation choice: time from home to work using train, car, or bus.

- GRETL doesn't include a routine to estimate conditional logit yet (as of version 1.8.1), so you'll want to use R to estimate this model

16.4.1 Conditional Logit Choice Probabilities

$$p_{ij} = P\left[\text{individual } i \text{ chooses alternative } j\right]$$

$$p_{ij} = \frac{\exp\left(\beta_{1j} + \beta_2 PRICE_{ij}\right)}{\exp\left(\beta_{11} + \beta_2 PRICE_{i1}\right) + \exp\left(\beta_{12} + \beta_2 PRICE_{i2}\right) + \exp\left(\beta_{13} + \beta_2 PRICE_{i3}\right)} \qquad (16.23)$$

16.4.1 Conditional Logit Choice Probabilities

$$P\left[y_{11}=1, y_{22}=1, y_{33}=1\right]= p_{11}\times p_{22}\times p_{33}$$

common

$$=\frac{\exp\left(\beta_{11}+\beta_{2}PRICE_{11}\right)}{\exp\left(\beta_{11}+\beta_{2}PRICE_{11}\right)+\exp\left(\beta_{12}+\beta_{2}PRICE_{12}\right)+\exp\left(\beta_{2}PRICE_{13}\right)}\times$$

$$\frac{\exp\left(\beta_{12}+\beta_{2}PRICE_{22}\right)}{\exp\left(\beta_{11}+\beta_{2}PRICE_{21}\right)+\exp\left(\beta_{12}+\beta_{2}PRICE_{22}\right)+\exp\left(\beta_{2}PRICE_{23}\right)}\times$$

$$\frac{\exp\left(\beta_{2}PRICE_{33}\right)}{\exp\left(\beta_{11}+\beta_{2}PRICE_{31}\right)+\exp\left(\beta_{12}+\beta_{2}PRICE_{32}\right)+\exp\left(\beta_{2}PRICE_{33}\right)}$$

We normalise one intercept to zero

$$=L\left(\beta_{12},\beta_{22},\beta_{2}\right)$$

16.4.2 Post-Estimation Analysis

- The own price effect is:

$$\frac{\partial p_{ij}}{\partial PRICE_{ij}} = p_{ij}\left(1 - p_{ij}\right)\beta_2 \qquad (16.24)$$

- The cross price effect is:

$$\frac{\partial p_{ij}}{\partial PRICE_{ik}} = -p_{ij}\, p_{ik}\beta_2 \qquad (16.25)$$

16.4.2 Post-Estimation Analysis

$$\frac{p_{ij}}{p_{ik}} = \frac{\exp\left(\beta_{1j} + \beta_2 PRICE_{ij}\right)}{\exp\left(\beta_{1k} + \beta_2 PRICE_{ik}\right)} = \exp\left[\left(\beta_{1j} - \beta_{1k}\right) + \beta_2\left(PRICE_{ij} - PRICE_{ik}\right)\right]$$

The odds ratio depends on the difference in prices, but not on the prices themselves. As in the multinomial logit model this ratio does not depend on the total number of alternatives, and there is the implicit assumption of the independence of irrelevant alternatives (IIA).

16.4.3 An Example

Table 16.4 Conditional Logit Parameter Estimates

Variables	Estimates	Standard errors	t-Statistics	p-Values
$PRICE(\beta_2)$	−2.2964	0.1377	−16.68	0.000
$PEPSI(\beta_{11})$	0.2832	0.0624	4.54	0.000
$7\text{-}UP(\beta_{12})$	0.1038	0.0625	1.66	0.096

16.4.3 An Example

The predicted probability of a Pepsi purchase, given that the price of Pepsi is $1, the price of 7-Up is $1.25 and the price of Coke is $1.10 is:

$$\hat{p}_{i1} = \frac{\exp\left(\tilde{\beta}_{11} + \tilde{\beta}_2 \times 1.00\right)}{\exp\left(\tilde{\beta}_{11} + \tilde{\beta}_2 \times 1.00\right) + \exp\left(\tilde{\beta}_{12} + \tilde{\beta}_2 \times 1.25\right) + \exp\left(\tilde{\beta}_2 \times 1.10\right)} = .4832$$

16.4.3 An Example

use http://www.stata-press.com/***data***/lf2/***travel2.dta***, ***clear***

```
. use http://www.stata-press.com/data/lf2/travel2.dta
(Greene & Hensher 1997 data on travel mode choice)

. list id mode train bus time invc choice in 1/6, sepby(id)
```

	id	mode	train	bus	time	invc	choice
1.	1	Train	1	0	406	31	0
2.	1	Bus	0	1	452	25	0
3.	1	Car	0	0	180	10	1
4.	2	Train	1	0	398	31	0
5.	2	Bus	0	1	452	25	0
6.	2	Car	0	0	255	11	1

An example

- For this transportation example, the dependent variable is ***choice***, a binary variable indicating which mode of transportation was chosen

- The regressors include the $J - 1$ dummy variables ***train*** and ***bus*** that identify each alternative mode of transportation and the alternative-specific variables ***time*** and ***invc*** **(**invc contains the in-vehicle cost of the trip: we expect that the higher the cost of traveling by some mode, the less likely a person is to choose that mode**)**

- Use the option group(id) to specify that the id variable identifies the groups in the sample

An example

Example from Greene and Hensher (1997) used by Long and Freese too illustrate ***clogit*** in STATA:

- Data on 152 groups (*id*) of travelers, choosing between three modes of travel: *train*, *bus* or *car*
- For each group, there are three rows of data corresponding to the three choices faced by each group, so we have $N \times J = 152 \times 3 = 456$ observations.

An example

- Two dummy variables (a third would be redundant) are used to indicate the mode of travel corresponding to a given row of data
- ***train*** is 1 if the observation has information about taking the train, else train is 0.
- ***bus*** is 1 if the observation contains information about taking a bus, else 0. If both train and bus are 0, the observation has information about driving a car.
- The actual choice made is shown by the dummy variable ***choice*** equal to 1 if the person took the mode of travel corresponding to a specific observation

An example

Estimates for *time* and *invc* are negative: the longer it takes to travel by a given mode, the less likely that mode is to be chosen. Similarly, the more it costs, the less likely a mode is to be chosen

```
. clogit choice train bus time invc, group(id) nolog

Conditional (fixed-effects) logistic regression   Number of obs   =        456
                                                  LR chi2(4)      =     172.06
                                                  Prob > chi2     =     0.0000
Log likelihood = -80.961135                       Pseudo R2       =     0.5152
```

choice	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
train	2.671238	.4531611	5.89	0.000	1.783058	3.559417
bus	1.472335	.4007152	3.67	0.000	.6869474	2.257722
time	-.0191453	.0024509	-7.81	0.000	-.0239489	-.0143417
invc	-.0481658	.0119516	-4.03	0.000	-.0715905	-.0247411

An example

```
. listcoef

clogit (N=456): Factor Change in Odds

  Odds of: 1 vs 0
```

choice	b	z	P>|z|	e^b
train	2.67124	5.895	0.000	14.4579
bus	1.47233	3.674	0.000	4.3594
time	-0.01915	-7.812	0.000	0.9810
invc	-0.04817	-4.030	0.000	0.9530

Odds-ratios

Everything else the same in time and invc, people prefer the bus and much prefer the train over the car

- For the alternative-specific variables, ***time*** and ***invc***, the odds ratios are the multiplicative effect of a unit change in a given independent variable on the odds of any given mode of travel
- E.g.: Increasing travel time by one minute for a given mode of transportation decreases the odds of using that mode of travel by a factor of .98 (2%), holding the values for the other alternatives constant
- If ***time*** for ***car*** increases in one minute while the time for ***train*** and ***bus*** remain the same, the odds of traveling by car decrease by 2 percent

- The odds ratios for the alternative-specific constants ***bus*** and ***train*** indicate the relative likelihood of choosing these options versus travelling by car (the base category), assuming that cost and time are the same for all options
- E.g.: If cost and time were equal, individuals would be 4.36 times more likely to travel by bus than by car, and they would be 14.46 times more likely to travel by train than by car

CLOGIT data structure

- Note that the data structure for the analysis of the conditional logit is rather special
- Long and Freese offer good advice on how to set up data that are structured in a more conventional fashion

CLOGIT vs MLOGIT

- Note that any multinomial logit model can be estimated using ***clogit*** by expanding the dataset (see Long and Freese for details) and respecifying the independent variables as a set of interactions
- This opens up the possibility of mixed models that include both individual-specific and alternative-specific variables (are richer travelers more likely to drive than to take the bus?)
- This opens up the possibility of imposing constraints on parameters in ***clogit*** that are not possible with ***mlogit*** (see Hendrickx 2001)

Keywords

- binary choice models
- censored data
- conditional logit
- count data models
- feasible generalized least squares
- Heckit
- identification problem
- independence of irrelevant alternatives (IIA)
- index models
- individual and alternative specific variables
- individual specific variables
- latent variables
- likelihood function
- limited dependent variables
- linear probability model
- logistic random variable
- logit
- log-likelihood function
- marginal effect
- maximum likelihood estimation
- multinomial choice models
- multinomial logit
- odds ratio
- ordered choice models
- ordered probit
- ordinal variables
- Poisson random variable
- Poisson regression model
- probit
- selection bias
- tobit model
- truncated data

References

- Long, S. and J. Freese for all topics (available on Google!)
- Cameron and Trivedi's book for count data

Next

- Ordered Choice
- Count data