2A: Multinomial outcomes: Basics

© A. Colin Cameron
U. of Calif. - Davis

.

OeNB Summer School 2010
Microeconometrics
Oesterreichische Nationalbank (OeNB), Vienna, Austria

Based on
A. Colin Cameron and Pravin K. Trivedi,
Microeconometrics: Methods and Applications (MMA), ch.15.
Microeconometrics using Stata (MUS), ch.15.
Data examples are from MUS.

Aug 30 - Sep 3, 2010

1. Introduction

- Multinomial outcome models: m possible discrete outcomes.
  - We model $\Pr[y = j|\mathbf{x}]$ for $j = 1, ..., m$.
- For cross-section data
  - Distribution for binary y is clearly multinomial.
  - Maximum likelihood estimator (MLE) is clearly best estimator.
  - It is fine to use default standard errors (robust is not needed).
- The main complications are:
  - Different models arise due to different specifications for $\Pr[y_i = j|\mathbf{x}_i]$
    - big distinction between unordered and ordered outcomes.
  - Interpretation of model estimates is complicated
    - separate ME for each outcome and possibly attribute for each outcome.

Outline

2. Multinomial data: Multinomial logit example

- Consider multinomial data on several mutually exclusive categorical outcomes.
- Example: multinomial variable y has outcome one of
  - $y = 1$ if fish from beach
  - $y = 2$ if fish from pier
  - $y = 3$ if fish from private boat
  - $y = 4$ if fish from charter boat
- Regressors are
  - price: varies by alternative and individual
  - catch rate: varies by alternative and individual
  - income: varies by individual but not alternative

- Variable definitions

```
. describe

Contains data from  mus15data.dta
  obs:         1,182
 vars:            16                          12 May 2008 20:46
 size:        85,104 (99.2% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
mode            float  %9.0g       modetype     Fishing mode
price           float  %9.0g                    price for chosen alternative
crate           float  %9.0g                    catch rate for chosen
                                                  alternative
dbeach          float  %9.0g                    1 if beach mode chosen
dpier           float  %9.0g                    1 if pier mode chosen
dprivate        float  %9.0g                    1 if private boat mode chosen
dcharter        float  %9.0g                    1 if charter boat mode chosen
pbeach          float  %9.0g                    price for beach mode
ppier           float  %9.0g                    price for pier mode
pprivate        float  %9.0g                    price for private boat mode
pcharter        float  %9.0g                    price for charter boat mode
qbeach          float  %9.0g                    catch rate for beach mode
qpier           float  %9.0g                    catch rate for pier mode
qprivate        float  %9.0g                    catch rate for private boat mode
qcharter        float  %9.0g                    catch rate for charter boat mode
income          float  %9.0g                    monthly income in thousands $
-------------------------------------------------------------------------------
```

- Data organization
  - here wide form with one observation per individual
  - each observation has data for all the possible alternatives.

```
. list mode d* p* income in 1/2, clean

          mode    dbeach    dpier   dprivate   dcharter     price    pbeach     ppier    pprivat
> e   pcharter   pmlogit1   pmlogit2   pmlogit3   pmlogit4     income
  1.   charter        0        0          0          1    182.93    157.93    157.93      157.9
> 3     182.93   .1125092   .0919656   .4516733   .3438518   7.083332
  2.   charter        0        0          0          1    34.534    15.114    15.114      10.53
> 4     34.534   .1122198   .2117394   .2635553   .4124855        1.25
```

- Here person 2 chose charter fishing (`mode=charter` or `dcharter=1`) when beach, pier, private and charter fishing cost, respectively, 15.11, 15.11, 10.53 and 34.53.

- Summary statistics
  - Columns $y = 1, ..., 4$ give sample means for those with $y = 1, ..., 4$.

Explanatory Variable	Sub-sample averages				
	y=1	y=2	y=3	y=4	All y
	Beach	Pier	Private	Charter	Overall
Income ($1,000's per month)	4.052	3.387	4.654	3.881	4.099
Price beach ($)	36	31	138	121	103
Price pier ($)	36	31	138	121	103
Price private ($)	98	82	42	45	55
Price charter ($)	125	110	71	75	84
Catch rate beach	0.28	0.26	0.21	0.25	0.24
Catch rate pier	0.22	0.20	0.13	0.16	0.16
Catch rate private	0.16	0.15	0.18	0.18	0.17
Catch rate charter	0.52	0.50	0.65	0.69	0.63
Sample probability	0.113	0.151	0.354	0.382	1.000
Observations	134	178	418	452	1182

- On average a person chooses to fish where it is cheapest to fish.

- Multinomial logit of fishing mode regressed on intercept and income
  - normalization that base outcome is beach fishing $(y = 1)$
  - $\Pr[y_{ij} = 1] = \frac{e^{\mathbf{x}_i'(\alpha_j + \beta_j \texttt{income})}}{\sum_{k=1}^{4} e^{\mathbf{x}_i'(\alpha_k + \beta_k \texttt{income})}}$, $j = 1, 2, 3, 4$, $\alpha_1 = 0$, $\beta_1 = 0$.

```
. * Multinomial logit with base outcome alternative 1
. mlogit mode income, baseoutcome(1)

Iteration 0:   log likelihood =    -1497.7229
Iteration 1:   log likelihood =    -1477.5265
Iteration 2:   log likelihood =    -1477.1514
Iteration 3:   log likelihood =    -1477.1506

Multinomial logistic regression                   Number of obs   =       1182
                                                   LR chi2( 3)     =      41.14
                                                   Prob > chi2     =     0.0000
Log likelihood =  -1477.1506                       Pseudo R2       =     0.0137
```

mode	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
pier						
income	-.1434029	.0532882	-2.69	0.007	-.2478459	-.03896
_cons	.8141503	.2286316	3.56	0.000	.3660405	1.26226
private						
income	.0919064	.0406638	2.26	0.024	.0122069	.1716059
_cons	.7389208	.1967309	3.76	0.000	.3533352	1.124506
charter						
income	-.0316399	.0418463	-0.76	0.450	-.1136571	.0503774
_cons	1.341291	.1945167	6.90	0.000	.9600457	1.722537

(mode==beach is the base outcome)

- Predicted probabilities of each outcome:

$$\widehat{\Pr}[y_{ij}=1]=\frac{e^{\mathbf{x}_i'(\widehat{\alpha}_j+\widehat{\beta}_j\texttt{income})}}{\sum_{k=1}^{4}e^{\mathbf{x}_i'(\widehat{\alpha}_k+\widehat{\beta}_k\texttt{income})}}$$

```
. * Compare average predicted probabilities to sample average frequencies
. predict pmlogit1 pmlogit2 pmlogit3 pmlogit4, pr

. summarize pmlogit* dbeach dpier dprivate dcharter, separator(4)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
pmlogit1	1182	.1133672	.0036716	.0947395	.1153659
pmlogit2	1182	.1505922	.0444575	.0356142	.2342903
pmlogit3	1182	.3536379	.0797714	.2396973	.625706
pmlogit4	1182	.3824027	.0346281	.2439403	.4158273
dbeach	1182	.1133672	.3171753	0	1
dpier	1182	.1505922	.3578023	0	1
dprivate	1182	.3536379	.4783008	0	1
dcharter	1182	.3824027	.4861799	0	1

- As expected average predicted probabilities sum to one.
- Furthermore average predicted probabilities of each outcome equals frequency of that outcome
  - Property of multinomial logit and conditional logit
  - Analog of OLS residuals sum to zero so $\overline{\widehat{y}}=\bar{y}$.

- Parameter interpretation is complex.
- Consider marginal effects - there are many (one for each outcome value)
  - Here $\text{ME}_{ij} = \partial p_{ij} / \partial \mathbf{x}_i = p_{ij}(\boldsymbol{\beta}_j - \overline{\boldsymbol{\beta}}_i)$ where $\overline{\boldsymbol{\beta}}_i = \sum_l p_{il} \boldsymbol{\beta}_l$.
  - e.g. marginal effect (AME and MEM) of $1,000 increase in annual income on probability fish from private boat (the third outcome).

```
. * MEM of income change for outcome 3
. mfx, predict(outcome(3))

Marginal effects after mlogit
      y  = Pr(mode==3) (predict, outcome(3))
         =  .35220366
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
income	.0325985	.00569	5.73	0.000	.021442	.043755	4.09934

```
.
. * AME of income change for outcome 3
. margeff, outcome(3)

Average partial effects after mlogit
      y  = Pr(mode=3)
```

variable	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
income	.0317562	.0052589	6.04	0.000	.021449	.0420633

- Stata 11 use `margins, dydx(*) predict(outcome(3))`

3. Multinomial data: Estimation Theory

- There are m mutually-exclusive alternatives.
  - Dependent variable y takes value j if the outcome is alternative j, $j = 1, ..., m$.
  - Probability that the outcome is alternative j is

$$p_j = \Pr[y = j], \qquad j = 1, ..., m.$$

- Introduce m binary variables for each observed y

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

  - Thus y_j equals 1 if alternative j is chosen and y_j equals 0 for all other non-chosen alternatives.
  - For an individual exactly one of $y_1, y_2, ..., y_m$ will be non-zero.
- The density for one observation can then be conveniently written as

$$f(y) = p_1^{y_1} \times p_2^{y_2} \times ... \times p_m^{y_m} = \prod_{j=1}^{m} p_j^{y_j}.$$

Estimation Theory: Multinomial regression models

- Probabilities depend on individual characteristics
  - parameterize p_{ij} in terms of observed data $\mathbf{x}_i$ and parameters $\boldsymbol{\beta}$:

$$p_{ij} = \Pr[y_i = j] = F_j(\mathbf{x}_i, \boldsymbol{\beta}), \quad j = 1, ..., m.$$

  - these probabilities should lie between 0 and 1 and sum over j to one.
- Examples
  - multinomial logit (with normalization $\boldsymbol{\beta}_1 = \mathbf{0}$)

$$p_{ij} = \Pr[y_i = j] = \frac{e^{\mathbf{x}_i' \boldsymbol{\beta}_j}}{\sum_{k=1}^{m} e^{\mathbf{x}_i' \boldsymbol{\beta}_k}}, \, j = 1, ..., m$$

  - other models (conditional logit, nested logit, etc.) use different p_{ij}.

- **Density** for one observation is

$$f(y_i|\mathbf{x}_i) = \prod_{j=1}^{m} p_{ij}^{y_{ij}} = \prod_{j=1}^{m} F_j(\mathbf{x}_i, \boldsymbol{\beta})^{y_{ij}}.$$

- **Likelihood** is product of densities given independence over i

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} f(y_i|\mathbf{x}_i) = \prod_{i=1}^{N} \prod_{j=1}^{m} p_{ij}^{y_{ij}}$$

- **Log-likelihood function**:

$$\begin{aligned} \ln \mathrm{L}(\boldsymbol{\beta}) &= \ln \left(\prod_{i=1}^{N} \prod_{j=1}^{m} p_{ij}^{y_{ij}} \right) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{m} y_{ij} \ln p_{ij} \\ &= \sum_{i=1}^{N} \sum_{j=1}^{m} y_{ij} \ln F_j(\mathbf{x}_i, \boldsymbol{\beta}). \end{aligned}$$

Estimation Theory: Maximum Likelihood Estimation

- **Log-likelihood function**:

$$\ln \mathrm{L}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \sum_{j=1}^{m} y_{ij} \ln p_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{m} y_{ij} \ln F_j(\mathbf{x}_i, \boldsymbol{\beta}).$$

- **MLE** maximizes $\ln \mathrm{L}(\boldsymbol{\beta})$ with first-order conditions:

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \sum_{j=1}^{m} \frac{y_{ij}}{F_j(\mathbf{x}_i, \boldsymbol{\beta})} \frac{\partial F_j(\mathbf{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}.$$

  - No explicit solution so use iterative gradient methods to compute $\widehat{\boldsymbol{\beta}}$.

Consistency of MLE

- What are weakest conditions for consistency?
  - The distribution is necessarily multinomial.
  - So for consistency need correct specification of $p_{ij} = F_j(\mathbf{x}_i, \boldsymbol{\beta})$.
  - Again qualitatively similar to linear model.
- $\widehat{\boldsymbol{\beta}}$ is asymptotically normal by the usual asymptotic theory if the d.g.p. is correctly specified.
  - For independent cross-section data there is no need to use robust se's.

4. Multinomial data: Types of Regressors

- Distinguish between two different types of regressors.
  - Alternative-specific or case-specific or alternative-invariant regressors do not vary across alternatives.
    - ⋆ e.g. income (in our example), gender.
  - Alternative-varying regressors may vary across alternatives.
    - ⋆ e.g. price.
- Models and commands can change
  - Multinomial logit: all regressors are case-specific.
  - Conditional logit: regressors are individual-varying.
- Required data format can change
  - wide form: N observations collapses data on all m alternatives
  - long form: $N \times m$ observations
  - Stata command `reshape` moves between long and wide forms.

5. Multinomial data: Multinomial logit model

- The multinomial logit (MNL) model is used when all regressors are individual-specific or case-specific.

$$p_{ij} = \Pr[y_i = j] = \frac{e^{\mathbf{x}_i' \boldsymbol{\beta}_j}}{\sum_{k=1}^{m} e^{\mathbf{x}_i' \boldsymbol{\beta}_k}}, \qquad j = 1, ..., m,$$

  - Clearly these probabilities lie between 0 and 1 and sum over j to one.
  - A normalization such as $\boldsymbol{\beta}_1 = \mathbf{0}$ is needed ensure model identification.
- The parameters $\boldsymbol{\beta}_2, ..., \boldsymbol{\beta}_m$ are estimated by MLE.
  - Some algebra gives f.o.c.

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k} = \sum_{i=1}^{N} (y_{ik} - p_{ik}) \mathbf{x}_i = \mathbf{0}, \quad k = 2, ..., m,$$

  - Clearly consistent if $\mathrm{E}[y_{ik} | \mathbf{x}_i] = p_{ik}$, i.e. correct model for probabilities.

Multinomial Logit: Marginal Effects

- For MNL (all case-specific) some algebra yields

$$\frac{\partial p_{ij}}{\partial \mathbf{x}_i} = p_{ij}(\boldsymbol{\beta}_j - \overline{\boldsymbol{\beta}}_i),$$

where $\overline{\boldsymbol{\beta}}_i = \sum_l p_{il}\boldsymbol{\beta}_l$ is a probability weighted average of the $\boldsymbol{\beta}_l$.

  - Coefficient signs $\neq$ sign of marginal effects (due to $\overline{\boldsymbol{\beta}}_i$)!!

- But can get sign if compare to base category (want good choice).
  - Suppose $\boldsymbol{\beta}_1 = \mathbf{0}$. Then for MNL it can be shown that

$$\Pr[y_i = j | y_i = j \text{ or } 1] = \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}_j}}{1 + e^{\mathbf{x}_i'\boldsymbol{\beta}_j}},$$

  - So $\widehat{\boldsymbol{\beta}}_j$ interpreted as parameters of a binary logit model between alternatives j and 1.

Multinomial Logit: Example

- Multinomial logit of fishing mode regressed on intercept and income
  - repeats earlier output
  - normalization that base outcome is beach fishing $(y = 1)$
  - so $\beta_{\text{priv,income}} = 0.0919$ means $\Pr[\text{priv}|\text{beach or priv}] \uparrow$ as income$\uparrow$

```
. * Multinomial logit with base outcome alternative 1
. mlogit mode income, baseoutcome(1)

Iteration 0:   log likelihood =-1497.7229
Iteration 1:   log likelihood =-1477.5265
Iteration 2:   log likelihood =-1477.1514
Iteration 3:   log likelihood =-1477.1506

Multinomial logistic regression                   Number of obs   =       1182
                                                  LR chi2(3)      =      41.14
                                                  Prob > chi2     =     0.0000
Log likelihood = -1477.1506                       Pseudo R2       =     0.0137
```

mode	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
pier						
income	-.1434029	.0532882	-2.69	0.007	-.2478459	-.03896
_cons	.8141503	.2286316	3.56	0.000	.3660405	1.26226
private						
income	.0919064	.0406638	2.26	0.024	.0122069	.1716059
_cons	.7389208	.1967309	3.76	0.000	.3533352	1.124506
charter						
income	-.0316399	.0418463	-0.76	0.450	-.1136571	.0503774
_cons	1.341291	.1945167	6.90	0.000	.9600457	1.722537

(mode==beach is the base outcome)

6. Multinomial data: Conditional logit model

- The conditional logit model (CL) is used when some or all regressors are alternative-specific.

$$p_{ij} = \Pr[y_i = j] = \frac{e^{\mathbf{x}_{ij}'\beta + \mathbf{z}_i'\gamma_j}}{\sum_{k=1}^{m} e^{\mathbf{x}_{ik}'\beta + \mathbf{z}_i'\gamma_k}}, \qquad j = 1, ..., m.$$

  - Here **x** are alternative-specific, **z** are case-specific regressors
  - A normalization such as $\gamma_1 = \mathbf{0}$ is needed for case-specific regressors.
  - Again these probabilities lie between 0 and 1 and sum over j to one.

- The parameters $\gamma_2, ..., \gamma_m$ are estimated by MLE.

Conditional Logit: Marginal Effects

- For CL some algebra yields

$$\frac{\partial p_{ij}}{\partial \mathbf{x}_{ik}} = \begin{cases} p_{ij}(1-p_{ij})\boldsymbol{\beta} & j=k \\ -p_{ij}p_{ik}\boldsymbol{\beta} & j \neq k \end{cases}$$

- It follows that if $\beta_j > 0$ then effect of x_{ij} increasing is
  - probability of outcome j increases
  - probabilities of all other outcomes decrease.
- Marginal effects in Stata 11
  - MEM and MER are computed using Stata command `estat mfx`
  - No command for AME.

Conditional Logit: Comparison to MNL

- The MNL and CL models are essentially the same model
  - MNL can always be re-expressed as CL (and vice-versa)
  - In particular, MNL can be estimated as a special case of CL.
- MNL formulation is often used in labor economics.
  - e.g. for choice of occupation all individual specific regressors, such as education, age and gender, are invariant across alternatives.
- CL formulation is more commonly used in transportation mode choice.
  - Then data is available on mode attributes such as price and time which vary over both individuals and alternatives.
  - When case-specific regressors are also included this is sometimes called a mixed model.

Conditional Logit: Example

- Conditional logit regression of fishing mode on both:
  - alternative-specific regressors: price and quality
  - case-specific regressors: intercept and income.
- First need to convert data from wide form to long form.
- Then estimate using command `asclogit`
  - This supplants older Stata command `clogit`.

```
* Original data are in wide form
use mus15data.dta, clear
generate id = _n
summarize id d* p* q* income
* Convert data from wide form to long form
reshape long d p q, i(id) j(fishmode beach pier ///
private charter) string
list in 1/4, clean noobs
summarize
* Conditional logit with alternative-specific and
case-specific regressors
asclogit d p q, case(id) alternatives(fishmode) ///
casevars(income) basealternative(beach) nolog
* Marg effect at mean of price change for each alternative
estat mfx, varlist(p)
* Predicted probabilities of choice of each mode
predict pasclogit, pr
```

```
. * Conditional logit with alternative-specific and case-specific regressors
. asclogit d p q, case(id) alternatives(fishmode) casevars(income) ///
>    basealternative(beach) nolog

Alternative-specific conditional logit         Number of obs      =      4728
Case variable: id                              Number of cases    =      1182

Alternative variable: fishmode                 Alts per case: min =         4
                                                              avg =       4.0
                                                              max =         4

                                                  Wald chi2(5)    =    252.98
Log likelihood = -1215.1376                       Prob > chi2     =    0.0000

------------------------------------------------------------------------------
           d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
fishmode     |
           p |  -.0251166   .0017317   -14.50   0.000    -.0285106   -.0217225
           q |    .357782   .1097733     3.26   0.001     .1426302    .5729337
-------------+----------------------------------------------------------------
beach        |  (base alternative)
-------------+----------------------------------------------------------------
charter      |
      income |  -.0332917   .0503409    -0.66   0.508     -.131958    .0653745
       _cons |   1.694366   .2240506     7.56   0.000     1.255235    2.133497
-------------+----------------------------------------------------------------
pier         |
      income |  -.1275771   .0506395    -2.52   0.012    -.2268288   -.0283255
       _cons |   .7779593   .2204939     3.53   0.000     .3457992    1.210119
-------------+----------------------------------------------------------------
private      |
      income |   .0894398   .0500671     1.79   0.074    -.0086898    .1875694
       _cons |   .5272788   .2227927     2.37   0.018     .0906132    .9639444
------------------------------------------------------------------------------
```

7. CL and MNL: Independence of Irrelevant Alternatives

- Consider CL model with all regressors case-specific:
  $p_j = e^{\mathbf{x}_j'\boldsymbol{\beta}} / \sum_l e^{\mathbf{x}_l'\boldsymbol{\beta}}$.
- Probability of observing alternative j given that either alternative j or alternative k is:

$$\begin{aligned}
\Pr[y=j|y=j \text{ or } k] &= \frac{p_j}{p_j+p_k} \\
&= \frac{e^{\mathbf{x}_j'\boldsymbol{\beta}} / \sum_l e^{\mathbf{x}_l'\boldsymbol{\beta}}}{\left(e^{\mathbf{x}_j'\boldsymbol{\beta}} / \sum_{l=1}^m e^{\mathbf{x}_l'\boldsymbol{\beta}}\right) + \left(e^{\mathbf{x}_k'\boldsymbol{\beta}} / \sum_{l=1}^m e^{\mathbf{x}_l'\boldsymbol{\beta}}\right)} \\
&= \frac{e^{\mathbf{x}_j'\boldsymbol{\beta}}}{e^{\mathbf{x}_j'\boldsymbol{\beta}} + e^{\mathbf{x}_k'\boldsymbol{\beta}}} \\
&= \frac{\exp((\mathbf{x}_j - \mathbf{x}_k)'\boldsymbol{\beta})}{1+\exp((\mathbf{x}_j - \mathbf{x}_k)'\boldsymbol{\beta})}.
\end{aligned}$$

- The choice between j and k is a binary logit model!!
  - The probability does not depend on other alternatives.
  - This restriction is called the assumption of independence of irrelevant alternatives (IIA).
  - The same result holds for the MNL model.

CL and MNL: Red Bus - Blue Bus problem

- In many consumer choice situations IIA assumption is too restrictive.
- Extreme example (red bus - blue bus problem)
  - In a CL or MNL model the conditional probability of commute by car given commute by car or red bus is independent of whether commuting by blue bus is an option.
  - But in practice we expect introduction of a blue bus, same as red bus in every aspect except color to
    - ★ have no impact on car use
    - ★ halve use of blue bus (bus riders split evenly between red and blue)
  - This leads to an increase in the conditional probability of car use given car or blue bus.
- This weakness of MNL and CL leads to extensions using a random utility approach in consumer choice applications.

8. Additive Random Utility Models (unordered models)

- For an m-alternative model the additive random utility model (ARUM) specifies utility of each alternative as

$$\begin{aligned} U_1 &= V_1 + \varepsilon_1 \\ U_2 &= V_2 + \varepsilon_2 \\ \vdots & \quad \vdots \quad \vdots \\ U_m &= V_m + \varepsilon_m \end{aligned}$$

- Here V_j is deterministic part of utility, e.g. $V_j = \mathbf{x}'\boldsymbol{\beta}_j$ or $\mathbf{x}_j'\boldsymbol{\beta}$, and ε_j are errors.
- Then j is chosen if it has the highest utility

$$\begin{aligned} \Pr[y=j] &= \Pr[U_j \geq U_k, \text{ all } k \neq j] \\ &= \Pr[\varepsilon_k - \varepsilon_j \leq -(V_k - V_j), \text{ all } k \neq j] \end{aligned}$$

Additive Random Utility Model: Examples

- Different error distributions lead to different multinomial models.
  - **1.** MNL and CL: ε_j are i.i.d. type I extreme value.
  - **2.** Nested logit: ε_j are correlated type I extreme value.
  - **3.** Random parameters logit: ε_j are i.i.d. type I extreme value but additionally parameters $\boldsymbol{\beta}_i$ are multivariate normal.
  - **4.** Multinomial probit: ε_j are multivariate normal.
- 1 and 2 are tractable.
  3 and 4 are not and require simulation methods to estimate.

Additive Random Utility Model: Three alternatives

- Choice probabilities can be complicated.
- For example, with three alternatives

$$\begin{aligned}
\Pr[y=3] &= \Pr[U_3 \geq U_1,\ U_3 \geq U_2] \\
&= \Pr[V_3+\varepsilon_3 \geq V_1+\varepsilon_1, \quad V_3+\varepsilon_3 \geq V_2+\varepsilon_2] \\
&= \Pr[\varepsilon_1-\varepsilon_3 \leq -(V_1-V_3), \quad \varepsilon_2-\varepsilon_3 \leq -(V_2-V_3)] \\
&= \Pr[\widetilde{\varepsilon}_{13} \leq -\widetilde{V}_{13}, \quad \widetilde{\varepsilon}_{23} \leq -\widetilde{V}_{23}] \\
&= \int_{-\infty}^{-\widetilde{V}_{13}} \int_{-\infty}^{-\widetilde{V}_{23}} f(\widetilde{\varepsilon}_{13}, \widetilde{\varepsilon}_{23})\, d\widetilde{\varepsilon}_{13}\, d\widetilde{\varepsilon}_{23}.
\end{aligned}$$

  - For a 3-alternative model probabilities require a bivariate integral.
- For an m-alternative model there is an $(m-1)$-variate integral.
- Methods have been developed to enable estimate models even if no analytical solution.

Additive Random Utility Model: MNL and CL

- Suppose the errors ε_j are i.i.d. type I extreme value distributed (also called log Weibull) with density

$$f(\varepsilon_j) = e^{-\varepsilon_j} \exp(e^{-\varepsilon_j}), \qquad j = 1, 2, ..., m.$$

- Then it can be shown that there is an analytical solution for probabilities

$$\begin{aligned} \Pr[y = j] &= \Pr[U_j \geq U_k, \text{ all } k \neq j] \\ &= \Pr[\varepsilon_k - \varepsilon_j \leq -(V_k - V_j), \text{ all } k \neq j] \\ &= \Pr[\widetilde{\varepsilon}_{kj} \leq -\widetilde{V}_{kj}] \\ &= \frac{e^{V_j}}{\sum_k e^{V_K}} \end{aligned}$$

  - MNL model is obtained when $V_j = \mathbf{x}'\boldsymbol{\beta}_j$
  - CL model is obtained when $V_j = \mathbf{x}_j'\boldsymbol{\beta}$.

ARUM: Independence of Irrelevant Alternatives revisited

- CL and MNL models arise when errors ε_j are independent across j.
- This is certain to be violated if two alternatives are similar:
  - Suppose alternatives 1 and 2 are similar
  - A low draw of ε_1 leads to overprediction of utility of alternative 1
  - But then expect to overpredict utility of alternative 1 $\Rightarrow$ ε_2 is low.
  - Low (or high) values of ε_1 and ε_2 $\Rightarrow$ errors are correlated.
- The earlier red bus - blue bus problem is an extreme case.
- So use richer models with correlated errors at the expense of increased computational burden: NL, RPL and MNP.

9. Multinomial data further models: unordered models

- Unordered model: no obvious ordering of alternatives.
- **1.** MNL and CL logit models
  - Easy to estimate
  - Too restrictive:
    - binary logit comparison between any pair of alternatives without consideration of other alternatives
    - red-bus blue-bus problem
- **2.** Nested logit
  - Easy to estimate
  - But need to specify nesting structure
- **3.** Random parameters logit:
  - Difficult to estimate but now feasible.
- **4.** Multinomial probit:
  - Difficult to estimate but now feasible.
  - Currently random parameters logit is preferred.

Multinomial data furtehr models: ordered models

- For outcomes for which there is a natural ordering
  - e.g. health status poor/fair ($y = 1$), good ($y = 2$), excellent ($y = 3$).
- Model is based on a single latent variable

$$y^* = \mathbf{x}'\boldsymbol{\beta} + u.$$

- Multinomial outcomes depend on magnitude of y^*. For 3 outcomes:

$$y_i = \begin{cases} 1 & \text{if } y^* \leq \alpha_1 \\ 2 & \text{if } \alpha_1 < y^* \leq \alpha_2 \\ 3 & \text{if } y^* > \alpha_2. \end{cases}$$

- MLE estimates $\boldsymbol{\beta}$ and the threshold parameters $\boldsymbol{\alpha}$
- Standard models are
  - Ordered probit model
  - Ordered logit model.

Multinomial data: model selection

- First choose between ordered and unordered.
- Within these if models are nested then can do a likelihood ratio test
  - CL and MNL are special case of nested logit
  - CL is special case of random parameters logit
- If models are not nonnested then use information criteria
  - These are log-likelihood with a degrees of freedom penalty.
    - $\text{AIC} = -2 \ln L + 2k$
    - $\text{BIC} = -2 \ln L + k \ln N$
  - Prefer model with smaller IC (as then larger $\ln L$)
  - BIC is preferred as AIC has too small a penalty for model size.
- Note: a likelihood ratio test of one restriction has critical value of $\chi^2_{.05}(1) = 3.84$
  - So LR tests favors more general model if $-2 \ln L$ falls by > 3.84
  - Also Vuong (1989) proposes LR test for nonnested models.

Multinomial data: Stata commands

- Stata commands

Command	Model
`mlogit`	multinomial logit
`asclogit`	conditional logit
`clogit`	older command for conditional logit
`nlogit`	nested logit (ARUM version)
`mprobit`	multinomial probit
`asmprobit`	multinomial probit
`mixlogit`	random parameters logit (Stata add-on)

- Commands `mlogit` and `mprobit` for individual-specific regressors only
  - data in wide form (one obs is all alternatives for individual)
- Other commands allow individual-varying regressors (e.g. price)
  - data in long form (one obs is one alternative for individual)
  - commands `reshape` to move from wide to long form.

10. Multinomial data: Ordered logit example

- Data example with health status poor/fair, good, or excellent.

```
. * Create multinomial ordered outcome variable with values y = 1, 2, 3
. use mus18data.dta, clear

. quietly keep if year==2

. generate hlthpf = hlthp + hlthf

. generate hlthe = (1 - hlthpf - hlthg)

. quietly generate hlthstat = 1 if hlthpf == 1

. quietly replace hlthstat = 2 if hlthg == 1

. quietly replace hlthstat = 3 if hlthe == 1

. label variable hlthstat "health status"

. label define hsvalue 1 poor_or_fair 2 good 3 excellent

. label values hlthstat hsvalue
```

- Summary statistics

```
. tabulate hlthstat

     health |
     status |      Freq.     Percent        Cum.
------------+-----------------------------------
poor_or_fair |        523        9.38        9.38
       good |      2,034       36.49       45.87
  excellent |      3,017       54.13      100.00
------------+-----------------------------------
      Total |      5,574      100.00

.
. * Summarize dependent and explanatory variables
. summarize hlthstat age linc ndisease

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    hlthstat |      5574    2.447435     .659524          1          3
         age |      5574    25.57613    16.73011   .0253251   63.27515
        linc |      5574    8.696929    1.220592          0   10.28324
    ndisease |      5574    11.20526    6.788959          0       58.6
```

- Estimate using Stata command ologit

```
. * Ordered logit estimates
. ologit hlthstat age linc ndisease, nolog

Ordered logistic regression                       Number of obs   =       5574
                                                  LR chi2(   3)     =     740.39
                                                  Prob > chi2     =     0.0000
Log likelihood =  -4769.8525                       Pseudo R2       =     0.0720
```

hlthstat	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	-.0292944	.001681	-17.43	0.000	-.0325891	-.0259996
linc	.2836537	.0231097	12.27	0.000	.2383594	.328948
ndisease	-.0549905	.0040692	-13.51	0.000	-.062966	-.047015
/cut1	-1.39598	.2061293			-1.799986	-.9919736
/cut2	.9513097	.2054294			.5486755	1.353944

- Compare fitted and actual frequencies and compute MEM
  - In Stata 11: `margins, predict(outcome(3)) dydx(*)` gives AME.

```
. summarize hlthpf hlthg hlthe p1ologit p2ologit p3ologit, separator(0)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      hlthpf |      5574    .0938285    .2916161          0          1
       hlthg |      5574    .3649085    .4814477          0          1
       hlthe |      5574     .541263    .4983392          0          1
    p1ologit |      5574    .0946903    .0843148   .0233629    .859022
    p2ologit |      5574    .3651672    .0946158   .1255265   .5276064
    p3ologit |      5574    .5401425    .1640575   .0154515   .7999009

.
. * Marginal effect at mean for 3rd outcome (health status excellent)
. mfx, predict(outcome(3))

Marginal effects after ologit
      y  = Pr(hlthstat==3) (predict, outcome(3))
         =   .53747616
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     age |  -.0072824      .00042  -17.43   0.000  -.008101 -.006463   25.5761
    linc |    .070515      .00575   12.26   0.000    .05924   .08179   8.69693
ndisease |  -.0136704      .00101  -13.50   0.000  -.015655 -.011686   11.2053
------------------------------------------------------------------------------
```

11. Some References

- The material is covered in graduate level texts including
  - CT(2005) MMA chapter 15 and CT(2009) MUS chapter 15
  - Wooldridge, J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.
  - Greene, W.H. (2007), *Econometric Analysis*, Prentice-Hall, Sixth edition.
- A classic book is
  - Maddala, G.S. (1986), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press.