Lecture 10: Multi-Choice Models

adapted from William Evans (Notre Dame) lecture note

March 18, 2025

Introduction

- In this section, we examine models with more than 2 possible choices
- Examples
  - How to get to work (bus, car, subway, walk)
  - How you treat a particular condition (bypass, heart cath, drugs, nothing)
  - Living arrangement (single, married, living with someone)
- In these examples, the choices reflect tradeoffs the consumer must face
  - Transportation: More flexibility usually requires more cost
  - Health: more invasive procedures may be more effective
- In contrast to ordered probit, no natural ordering of choices

Modeling Choices

- Model is designed to estimate what cofactors predict choice of 1 from the other J-1 alternatives
- Motivated from the same decision/theoretic perspective used in logit/probit modes
- Just have expanded the choice set

Some Model Specifics

- j indexes choices (J of them)

  No need to assume equal choices
- i indexes people (N of them)
- $Y_{ij} = 1$ if person i selects option $j, = 0$ otherwise
- U_{ij} is the utility or net benefit of person "i" if they select option "j"
- Suppose they select option 1

Some Model Specifics

- Then there are a set of $(J-1)$ inequalities that must be true

$$\begin{aligned} &U_{i1} > U_{i2} \\ &U_{i1} > U_{i3} \\ &\cdots \\ &U_{i1} > U_{iJ} \end{aligned}$$

- Choice 1 dominates the other
- We will use the $(J-1)$ inequality to help build the model

Two Different but Similar Models

- Multinomial logit
  - Utility varies only by "i" characteristics
  - People of different incomes more likely to pick one mode of transportation
- Conditional logit
  - Utility varies only by the characteristics of the option
  - Each mode of transportation has different costs/time
- Mixed logit - combined the two

Multinomial Logit

- Utility is determined by two parts: observed and unobserved characteristics (just like logit)
- However, measured components only vary at the individual level
- Therefore, the model measures what characteristics predict choice

  Are people of different income levels more/less likely to take one mode of transportation to work

Multinomial Logit

- $U_{ij} = X_i \beta_j + \varepsilon_{ij}$
- ε_{ij} is assumed to be a type 1 extreme value distribution

$$f(\varepsilon_{ij}) = \exp(-\varepsilon_{ij}) \exp(-\exp(-\varepsilon_{ij}))$$
$$F(a) = \exp(-\exp(-a))$$

- Choice of 1 implies utility from 1 exceeds that of options 2 (and 3 and 4....)

Multinomial Logit

- Focus on choice of option 1 first
- $U_{i1} > U_{i2}$ implies that

$$X_i\beta_1 + \varepsilon_{i1} > X_i\beta_2 + \varepsilon_{i2}$$

OR

$$\varepsilon_{i2} < X_i\beta_1 - X_i\beta_2 + \varepsilon_{i1}$$

Multinomial Logit

- There is $(J-1)$ of these inequalities

$$\begin{aligned}
&\varepsilon_{i2} < X_i\beta_1 - X_i\beta_2 + \varepsilon_{i1} \\
&\varepsilon_{i3} < X_i\beta_1 - X_i\beta_3 + \varepsilon_{i1} \\
&\ldots \\
&\varepsilon_{iJ} < X_i\beta_1 - X_i\beta_j + \varepsilon_{i1}
\end{aligned}$$

- Probability we observe option 1 selected is therefore

$$\begin{gathered}
\text{Prob}(\varepsilon_{i2} < X_i\beta_1 - X_i\beta_2 + \varepsilon_{i1} \cap \varepsilon_{i3} < X_i\beta_1 - X_i\beta_3 + \varepsilon_{i1} \\
\ldots \cap \varepsilon_{iJ} < X_i\beta_1 - X_i\beta_j + \varepsilon_{i1})
\end{gathered}$$

Multinomial Logit

- Recall: if a, b and c are independent
- $\Pr(A \cap B \cap C) = \Pr(A)\Pr(B)\Pr(C)$
- And since $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_J$ are independent
- The probability of observing 1 selected equals
- $\Pr\left(X_i\beta_1 - X_i\beta_2 + \varepsilon_{i1}\right)\Pr\left(X_i\beta_1 - X_i\beta_3 + \varepsilon_{i1}\right)\ldots$
- But since ε_1 is a random variable, must integrate this value out

$$\int_{-\infty}^{\infty} \prod_{j=2}^{J} F\left(\varepsilon_{1i} + X_i\beta_1 - X_i\beta_j\right) f\left(\varepsilon_{1i}\right) d\varepsilon_{1i}$$

$$= \frac{\exp\left(X_i\beta_1\right)}{\sum_{j=1}^{J} \exp\left(X_i\beta_j\right)}$$

General Result

- The probability you choose option j is

$$\text{Prob}\left(Y_{ij} = 1 \mid X_i\right) = \exp\left(X_i \beta_j\right) / \Sigma_k \left[\exp\left(X_{ik} \beta_k\right)\right]$$

- Each option j has a different vector β_j
- To identify the model, must pick one option (m) as the "base" or "reference" option and set $\beta_m = 0$
- Therefore, the coefficients for β_j represent the impact of a personal characteristic on the option they will select j relative to m.
- If $J = 2$, model collapses to logit

General Result

- Log likelihood function
- $Y_{ij} = 1$ of person I chose option j
  0 otherwise
- $\text{Prob}\,(Y_{ij} = 1)$ is the estimated probability option j will be picked

$$L = \sum_i \sum_j Y_{ij} \ln\left[\text{Prob}\,(Y_{ij})\right]$$

Estimating in STATA

- Estimation is trivial so long as data is constructed properly
- Suppose individuals are making the decision. There is one observation per person
- The observation must identify
  - the X's
  - the options selected

Estimating in STATA

- 1500 adult females who were part of a job training program
- They enrolled in one of 4 job training programs
- Choice identifies what option was picked
  - 1 = classroom training
  - 2 = on the job training
  - 3 = job search assistance
  - 4 = other

Estimating in STATA

- `* get frequency of choice variable;`
- `. tab choice;`

```
      choice |      Freq.     Percent        Cum.
-------------+-----------------------------------
           1 |        642       42.80       42.80
           2 |        225       15.00       57.80
           3 |        331       22.07       79.87
           4 |        302       20.13      100.00
-------------+-----------------------------------
       Total |      1,500      100.00
```

Estimating in STATA

- Syntax of *mlogit* procedure. Identical to logit but, must list as an option the choice to be used as the reference (base) option
- *Mlogit dep.var ind.var, base(#)*
- Example from program
- *mlogit choice age black hisp nvrwrk lths hsgrad, base*(4)

Estimating in STATA

- Sets of characteristics are used to explain what option was picked
  - Age
  - Race/ethnicity
  - Education
  - Whether respondent worked in the past
- 1500 obsservations in the data set

Estimating in STATA

```
•   Multinomial logistic regression                       Number of obs   =       1500
•                                                          LR chi2(18)     =     135.19
•                                                          Prob > chi2     =     0.0000
•   Log likelihood = -1888.2957                            Pseudo R2       =     0.0346

•   ------------------------------------------------------------------------------
•         choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
•   -------------+----------------------------------------------------------------
•   1            |
•            age |   .0071385   .0081098     0.88   0.379    -.0087564    .0230334
•          black |   1.219628   .1833561     6.65   0.000     .8602566    1.578999
•           hisp |   .0372041   .2238755     0.17   0.868    -.4015838     .475992
•         nvrwrk |   .0747461    .190311     0.39   0.694    -.2982567    .4477489
•           lths |  -.0084065   .2065292    -0.04   0.968    -.4131964    .3963833
•         hsgrad |   .3780081   .2079569     1.82   0.069    -.0295799     .785596
•          _cons |   .0295614   .3287135     0.09   0.928    -.6147052    .6738279
•   -------------+----------------------------------------------------------------
```

Estimating in STATA

```
------------+----------------------------------------------------------------
2           |
        age |    .008348   .0099828     0.84   0.403     -.011218    .0279139
      black |   .5236467   .2263064     2.31   0.021     .0800942    .9671992
       hisp |  -.8671109   .3589538    -2.42   0.016    -1.570647   -.1635743
     nvrwrk |   -.704571   .2840205    -2.48   0.013    -1.261241   -.1479011
       lths |  -.3472458   .2454952    -1.41   0.157    -.8284075    .1339159
     hsgrad |  -.0812244   .2454501    -0.33   0.741    -.5622979     .399849
      _cons |  -.3362433   .3981894    -0.84   0.398     -1.11668    .4441936
------------+----------------------------------------------------------------
3           |
        age |    .030957   .0087291     3.55   0.000     .0138483    .0480657
      black |    .835996   .2102365     3.98   0.000     .4239399    1.248052
       hisp |   .5933104   .2372465     2.50   0.012     .1283157    1.058305
     nvrwrk |  -.6829221   .2432276    -2.81   0.005    -1.159639   -.2062047
       lths |  -.4399217   .2281054    -1.93   0.054        -.887    .0071566
     hsgrad |   .1041374   .2248972     0.46   0.643    -.3366529    .5449278
      _cons |  -.9863286   .3613369    -2.73   0.006    -1.694536   -.2781213
-----------------------------------------------------------------------------
```

Estimating in STATA

- Notice there is a separate constant for each alternative
- Represents that, given X's, some options are more popular than others
- Constants measure in reference to the base alternative

How to interpret parameters

- Parameters in and of themselves not that informative
- We want to know how the probabilities of picking one option will change if we change X
- Two types of X' s
  - Continuous
  - dichotomous

How to interpret parameters

- Probability of choosing option j

$$\text{Prob}\left(Y_{ij}=1 \mid X_i\right)=\exp\left(X_i\beta_j\right)/\Sigma_k\left[\exp\left(X_i\beta_k\right)\right]$$

- $X_i=\left(X_{i1},X_{i2},\ldots.X_{ik}\right)$
- Suppose $X_{\mathrm{i}1}$ is continuous
- $d\,\text{Prob}\left(Y_{ij}=1 \mid X_j\right)/dX_{i1}=?$

Suppose X_{i1} is Continuous

- Calculate the marginal effect

$$\text{dProb}\left(Y_{ij} = 1 \mid X_i\right)/dX_{i1}$$

  where $\mathrm{X_i}$ is evaluated at the sample means

- Can show that

$$\text{dProb}\left(Y_{ij} = 1 \mid X_i\right)/dX_{i1} = P_j\left[\beta_{1j} - b\right]$$

  Where $b = P_1\beta_{11} + P_2\beta_{12} + \ldots . P_k\beta_{1k}$

Suppose X_{i1} is Continuous

- The marginal effect is the difference in the parameter for option 1 and a weighted average of all the parameters on the 1^{st} variable
- Weights are the initial probabilities of picking the option
- Notice that the 'sign' of beta does not inform you about the sign of the ME

Suppose X_{i2} is Dichotomous

- Calculate change in probabilities

$$P_1 = \text{Prob}\left(Y_{ij} = 1 \mid X_{i1}, X_{i2} = 1 \ldots . X_{ik}\right)$$

$$P_0 = \text{Prob}\left(Y_{ij} = 1 \mid X_{i1}, X_{i2} = 0 \ldots . X_{ik}\right)$$

- $ATE = P_1 - P_0$
- Stata uses sample means for the X's

β and Marginal Effects

	Option 1		Option 2		Option 3	
	β	ME	β	ME	β	ME
Age	0.007	-0.002	0.008	-0.001	0.031	0.004
Black	1.219	0.179	0.524	-0.042	0.836	0.001
Hisp	0.037	-0.020	-0.867	-0.100	0.593	0.136
Nvrwk	0.075	0.121	-0.704	-0.065	-0.682	-0.093
LTHS	-0.008	0.065	-0.347	-0.029	-0.449	-0.062
HS	0.378	0.088	-0.336	-0.038	0.104	-0.016

β and Marginal Effects

- Notice that there is not a direct correspondence between sign of β and the sign of the marginal effect
- Really need to calculate the ME' s to know what is going on

Problem: IIA

- Independent of Irrelevant alternatives or 'red bus/blue bus' problem
- Suppose two options to get to work
  - Car (option c)
  - Blue bus (option b)
- What are the odds of choosing option c over b?

Problem: IIA

- Since numerator is the same in all probabilities

$$\Pr(Y_{ic} = 1 \mid X_i) / \Pr(Y_{ib} = 1 \mid X_i) = \exp(X_i\beta_c) / \exp(X_i\beta_b)$$

- Note two thing: Odds are
  - independent of the number of alternatives
  - Independent of characteristics of alt.
  - Not appealing

Example

- Pr(Car) + Pr(Bus) = 1 (by definition)
- Originally, let's assume
  - Pr(Car) = 0.75
  - Pr(Blue Bus) = 0.25
- So odds of picking the car is 3

Example

- Suppose that the local govt. introduces a new bus.
- Identical in every way to old bus but it is now red (option r)
- Choice set has expanded but not improved
  - Commuters should not be any more likely to ride a bus because it is red
  - Should not decrease the chance you take the car
- In reality, red bus should just cut into the blue bus business
  - Pr(Car) = 0.75
  - Pr(Red Bus) = 0.125 = Pr(Blue Bus)
  - Odds of taking car/blue bus = 6

What does model suggest?

- Since red/blue bus are identical $\beta_{\mathrm{b}} = \beta_{\mathrm{r}}$
- Therefore,

$$\Pr\left(Y_{ib} = 1 \mid X_i\right) / \Pr\left(Y_{ir} = 1 \mid X_i\right) = \exp(X_i\beta_b) / \exp\left(X_i\beta_r\right) = 1$$

- But, because the odds are independent of other alternatives

$$\Pr\left(Y_{ic} = 1 \mid X_j\right) / \Pr\left(Y_{ib} = 1 \mid X_j\right) = \exp\left(X_i\beta_c\right) / \exp\left(X_i\beta_b\right) = 3 \text{ still}$$

What does model suggest?

- With these new odds, then
  - Pr(Car) = 0.6
  - Pr(Blue) = 0.2
  - Pr(Red) = 0.2
- Note the model predicts a large decline in car traffic even though the person has not been made better off by the introduction of the new option
- Poorly labeled - really independence of relevant alternatives
- Implication? When you use these models to simulate what will happen if a new alternative is added, will predict much larger changes than will happen

How to Get Around IIA

- Conditional probit models.
  - Allow for correlation in errors
  - Very complicated.
  - Not pre-programmed into any statistical package
- Nested logit
  - Group choices into similar categories
  - IIA within category and between category

How to Get Around IIA

- Example: Model of car choice
  - 4 options: Sedan, minivan, SUV, pickup truck
- Could nest the decision
- First decide whether you want something on a car or truck platform
- Then pick with the group
  - Car: sedan or minivan
  - Truck: pickup or SUV

How to Get Around IIA

- IIA is imposed
- Within a nest:
  - Cars/minivans
  - Pickup and SUV
- Between 1^{st} level decision
  - Truck and car platform

Conditional Logit

- Devised by McFadden and similar to logit
- Allows characteristics to vary across alternatives
- $U_{ij} = Z_{ij}\gamma + \varepsilon_{ij}$
- ε_{ij} is again assumed to be a type 1 extreme value distribution

Conditional Logit

- Choice of 1 over $2, 3, \ldots J$ generates $J - 1$ inequalities
- Reduces to similar probability as before
- Probability of choosing option j

$$\text{Prob}\left(Y_{ij} = 1 \mid Z_{ij}\right) = \exp\left(Z_{ij}\gamma\right) / \Sigma_k \left[\exp\left(Z_{ik}\gamma\right)\right]$$

Mixed models

- Most frequent type of multiple unordered choice
- Z's that vary by option
- X's that vary by person

$$U_{ij} = X_i \beta_j + Z_{ij} \gamma + \varepsilon_{ij}$$

$$\text{Prob}\left(Y_{ij} = 1 \mid X_i Z_{ij}\right) = \exp\left(X_i \beta_j + Z_{ij} \gamma\right) / \Sigma_k \left[\exp\left(X_i \beta_k + Z_{ik} \gamma\right)\right]$$

How must data be structured?

- There must be J observations (one for each alternative) for each person (N) in the data set

  NJ observations in total
- Must be an ID variable that identifies what observations go together
- A *dummy variable* that equals 1 identifies the observation from the J alternatives that is selected

How must data be structured?

- Example

  Travel_choice_example.dta

- 210 families had one of four ways to travel to another city in Australia

  Fly (mode = 1)

  Train (= 2)

  Bus (= 3)

  Car (= 4)

- Two variables that vary by option/person

  Costs and travel time

- One family-specific characteristic – Income

How must data be structured?

Household Index

Index of Options

Actual Choice

Travel time In minutes

Travel cost In $

1005	1	0	208	82	45	2
1005	2	0	448	93	45	2
1005	3	0	502	94	45	2
1005	4	1	600	99	45	2
1006	1	0	169	70	20	1
1006	2	1	385	57	20	1
1006	3	0	452	58	20	1
1006	4	0	284	43	20	1

Household income X 1000

Size of group traveling

48

Preparing the Data for Estimation

- There are 4 choices. Some more likely than others.
- Need to reflect this by having $J - 1$ dummy variables
- Construct dummies for air, bus, train choices

  $gen\ air{=}mode == 1$

  $gen\ train = mode == 2$

  gen bus=mode==3

Preparing the Data for Estimation

For each family-specific characteristic, need to *interact with* a option dummy variable

```
* interact hhinc with choice dummies;
• gen hhinc_air=air*hhinc;
• gen hhinc_train=train*hhinc;
• gen hhinc_bus=bus*hhinc;
```

Preparing the Data for Estimation

Costs are a little complicated

- If by car, costs are costs.
- If by air/bus/train, costs are groupsize*costs (need to buy a ticket for all travelers)

```
gen group_costs=car*costs
    +(1-car)*groupsize*costs;
```

Preparing the Data for Estimation

1=air, 2=train, 3=bus, 4=car	=1 if choice, =0 otherwise		
	0	1	Total
1	152	58	210
2	147	63	210
3	180	30	210
4	151	59	210
Total	630	210	840

Means of Variables

	% selecting	Costs	Travel time
Plane	27.6%	$174	194
Train	30.0%	$237	583
Bus	14.3%	$212	671
Car	28.1%	$95	573

Stata Codes

Run two models. One with only variables that vary by option (conditional logit)

```
clogit choice air train bus time
totalcosts, group(hhid);
```

Run another with family characteristics

```
clogit choice air train bus time
totalcosts hhinc_*, group(hhid);
```

Results from Second Model

```
Conditional (fixed-effects) logistic regression   Number of obs   =        840
                                                  LR chi2(8)      =     102.15
                                                  Prob > chi2     =     0.0000
Log likelihood = -240.04567                       Pseudo R2       =     0.1754
------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         air |  -1.393948   .6314865    -2.21   0.027    -2.631639   -.1562576
       train |   2.371822   .4460489     5.32   0.000     1.497582    3.246062
         bus |   1.147733   .5159572     2.22   0.026     .1364751     2.15899
        time |  -.0036407   .0007603    -4.79   0.000    -.0051308   -.0021506
 group_costs |  -.0036817   .0013058    -2.82   0.005    -.0062411   -.0011224
   hhinc_air |   .0058589   .0106655     0.55   0.583    -.0150451     .026763
 hhinc_train |  -.0492424   .0119151    -4.13   0.000    -.0725956   -.0258892
   hhinc_bus |  -.0290673   .0131363    -2.21   0.027    -.0548141   -.0033206
------------------------------------------------------------------------------
```

Problem

- The post-estimation subrountines have not been written for CLOGIT
- Need to brute force the outcomes
- On next slide, some code to estimate change in probabilities if travel time by car increases by 30 minutes

Codes

- `predict pred0;`
- `replace time=time+30 if mode==4;`
- `predict pred30;`
- `gen change_p=pred30-pred0;`

- `sum change_p if mode==1;`
- `sum change_p if mode==2;`
- `sum change_p if mode==3;`
- `sum change_p if mode==4;`

Results

Change in probabilities

$$\left.\begin{array}{lr} \text{Air} & 0.0083 \\ \text{Train} & 0.0067 \\ \text{Bus} & 0.0037 \\ \text{Car} & -0.0187 \end{array}\right\} 0.0187$$