Conditional Logit

Categorical and Limited
Dependent Variables
Paul A. Jargowsky

Choice Characteristics

- Income is a characteristic of the person.
- Cost is a characteristic of the choice.
- Costs of the alternatives can vary by person, but each person has a cost for each alternative.
- The multinomial model as developed earlier only allows for individual characteristics.
- Surely choice characteristics should also matter!
- Conditional logit, developed next, allows you to analyze choices conditional on the choice characteristics.
- Later, we combine choice and individual characteristics.

Choice Characteristics

$$U_{1i} = \alpha_1 + \alpha_2 C_{1i} + \varepsilon_{1i}$$
$$U_{2i} = \alpha_1 + \alpha_2 C_{2i} + \varepsilon_{2i}$$

Notice the subscripts. How are they different from multinomial logit?

$$\begin{aligned} P_1 &= \Pr\left(Y_i = 1 \mid C_{1i}, C_{2i}\right) \\ &= \Pr\left[\left(\alpha_1 + \alpha_2 C_{1i} + \varepsilon_{1i}\right) > \left(\alpha_1 + \alpha_2 C_{2i} + \varepsilon_{2i}\right)\right] \\ &= \Pr\left[\left(\varepsilon_{1i} - \varepsilon_{2i}\right) > \left(\alpha_1 + \alpha_2 C_{2i} - \alpha_1 - \alpha_2 C_{1i}\right)\right] \\ &= \Pr\left[\left(\varepsilon_{1i} - \varepsilon_{2i}\right) > \alpha_2\left(C_{2i} - C_{1i}\right)\right] \end{aligned}$$

Question: is the model identified? Why or why not?

Yes, the model is identified, except the constant.
The C's are known, and only α_2 needs to be estimated.

Generalizing

$$\begin{aligned} U_{1i} &= \alpha_1 + \alpha_2 C_{1i} + \alpha_3 T_{1i} + \ldots + \varepsilon_{1i} = \mathbf{z}_{1i}\boldsymbol{\alpha} + \varepsilon_{1i} \\ U_{2i} &= \alpha_1 + \alpha_2 C_{2i} + \alpha_3 T_{2i} + \ldots + \varepsilon_{2i} = \mathbf{z}_{2i}\boldsymbol{\alpha} + \varepsilon_{2i} \\ &\vdots \\ U_{Ji} &= \alpha_1 + \alpha_2 C_{Ji} + \alpha_3 T_{Ji} + \ldots + \varepsilon_{Ji} = \mathbf{z}_{Ji}\boldsymbol{\alpha} + \varepsilon_{Ji} \end{aligned}$$

Assuming IID Type I extreme value distributions for the ε's leads to the following form for the probability of outcome 1, and the other are analogous.

$$P_{1i} = \frac{e^{\mathbf{z}_{1i}\boldsymbol{\alpha}}}{\sum_{j=1}^{J} e^{\mathbf{z}_{ji}\boldsymbol{\alpha}}}$$

From here, construct the likelihood function and use MLE.

Creating a New Data Structure

```
person      income    pcar   pbus   pwalk   choice
     1    11.26115    7.05   1.40    0.00        3

 * transform data to person/choice records
 expand 3
 sort person
 by person: gen rectype = _n

 gen price = pcar if rectype==1
 replace price = pbus if rectype==2
 replace price = pwalk if rectype==3

 gen picked = (choice==rectype)
```

person	*_n*
1	1
1	2
1	3
2	1
2	2
2	3
3	1
⋮	⋮

```
person      income    pcar   pbus   pwalk   choice   rectype   price   picked
     1    11.26115    7.05   1.40    0.00        3         1    7.05        0
     1    11.26115    7.05   1.40    0.00        3         2    1.40        0
     1    11.26115    7.05   1.40    0.00        3         3    0.00        1
```

Now there is one record for each person/alternative combination.

Stata: asclogit

. asclogit picked price, case(person) alternatives(rectype) nocons

```
Alternative-specific conditional logit         Number of obs      =       3000
Case variable: person                          Number of cases    =       1000

Alternative variable: rectype                  Alts per case: min =          3
                                                              avg =        3.0
                                                              max =          3

                                                  Wald chi2(1)    =       4.76
Log likelihood = -1096.2518                       Prob > chi2     =     0.0291

------------------------------------------------------------------------------
      picked |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rectype      |
       price |   .0227412    .010419     2.18   0.029     .0023202    .0431621
------------------------------------------------------------------------------
```

Predictions for Person 1

$$\hat{U}_{1i} = \hat{\alpha}C_{1i} = .02274(7.05) = 0.1603$$
$$\hat{U}_{2i} = \hat{\alpha}C_{2i} = .02274(1.40) = 0.0318$$
$$\hat{U}_{3i} = \hat{\alpha}C_{3i} = .02274(0) = 0$$

$$\hat{P}_2 = \frac{1.032}{3.206} = 0.322 \qquad \hat{P}_3 = \frac{1}{3.206} = 0.312$$

```
. predict prob

. list choice price  picked prob if person==1

     +---------------------------------------+
     | choice      price   picked       prob |
     |---------------------------------------|
  1. |   Walk   7.050649        0   .3661291 |
  2. |   Walk   1.400565        0   .3219823 |
  3. |   Walk          0        1   .3118886 |
     +---------------------------------------+
```

$$\hat{P}_1 = \frac{e^{\hat{U}_1}}{e^{\hat{U}_1} + e^{\hat{U}_2} + e^{\hat{U}_3}} = \frac{e^{0.1603}}{e^{0.1603} + e^{0.0318} + e^{0}} = \frac{1.174}{1.174 + 1.032 + 1} = \frac{1.174}{3.206} = 0.366$$

What if all alternatives had the same price?

$$\hat{U}_{1i} = \hat{\alpha}C_{1i} = .02274(1) = 0.02274$$
$$\hat{U}_{2i} = \hat{\alpha}C_{2i} = .02274(1) = 0.02274$$
$$\hat{U}_{3i} = \hat{\alpha}C_{3i} = .02274(1) = 0.02274$$

$$\hat{P}_2 = \frac{1.032}{3.206} = 0.322 \qquad \hat{P}_3 = \frac{1}{3.206} = 0.312$$

$$\hat{P}_1 = \frac{e^{\hat{U}_1}}{e^{\hat{U}_1} + e^{\hat{U}_2} + e^{\hat{U}_3}} = \frac{e^{0.02274}}{e^{0.02274} + e^{0.02274} + e^{0.02274}} = \frac{1.023}{1.023 + 1.023 + 1.023} = \frac{1}{3}$$

Implies that the probability for any option depends only on price, and that if two alternatives have the same price, they have the same probability. Not relatistic! We will fix this later.

```
. estat mfx

Pr(choice = 1|1 selected) = .36430563
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
price        |
           1 |  .005267   .002502    2.10   0.035   .000363   .010171    6.9902
           2 | -.002692   .001306   -2.06   0.039  -.005251  -.000133    1.9607
           3 | -.002575   .001196   -2.15   0.031  -.004919   -.00023         0

Pr(choice = 2|1 selected) = .32493211
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
price        |
           1 | -.002692   .001306   -2.06   0.039  -.005251  -.000133    6.9902
           2 |  .004988   .002252    2.21   0.027   .000574   .009403    1.9607
           3 | -.002296   .000947   -2.43   0.015  -.004152  -.000441         0

Pr(choice = 3|1 selected) = .31076226
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
price        |
           1 | -.002575   .001196   -2.15   0.031  -.004919   -.00023    6.9902
           2 | -.002296   .000947   -2.43   0.015  -.004152  -.000441    1.9607
           3 |  .004871   .002143    2.27   0.023   .000671   .009071         0
```

```
. estat mfx, at(1:price=5 2:price=2 3:price=0)

Pr(choice = 1|1 selected) =  .3537854
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
price        |
           1 |  .005199   .002445    2.13   0.033   .000407   .009991         5
           2 | -.002659   .001277   -2.08   0.037  -.005162  -.000155         2
           3 |  -.00254   .001168   -2.18   0.030  -.004829  -.000252         0

Pr(choice = 2|1 selected) = .33045387
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
price        |
           1 | -.002659   .001277   -2.08   0.037  -.005162  -.000155         5
           2 |  .005032   .002294    2.19   0.028   .000536   .009527         2
           3 | -.002373   .001016   -2.33   0.020  -.004365  -.000381         0

Pr(choice = 3|1 selected) = .31576073
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
price        |
           1 |  -.00254   .001168   -2.18   0.030  -.004829  -.000252         5
           2 | -.002373   .001016   -2.33   0.020  -.004365  -.000381         2
           3 |  .004913   .002184    2.25   0.024   .000633   .009194         0
```

Odds of Choice M vs. N

$$P_{mi} = \frac{e^{\mathbf{z}_{mi}\boldsymbol{\alpha}}}{\sum_{j=1}^{J} e^{\mathbf{z}_{ji}\boldsymbol{\alpha}}} \qquad P_{ni} = \frac{e^{\mathbf{z}_{ni}\boldsymbol{\alpha}}}{\sum_{j=1}^{J} e^{\mathbf{z}_{ji}\boldsymbol{\alpha}}}$$

Let $\mathbf{z}_{ji}\boldsymbol{\alpha} = \alpha C_{ji}$

$$\Omega_{m|n} = \frac{P_{mi}}{P_{ni}} = \frac{e^{\alpha C_{mi}}}{e^{\alpha C_{ni}}} = e^{\alpha C_{mi} - \alpha C_{ni}}$$

Odds of m vs. n.

$$\Omega'_{m|n} = \frac{P_{mi}}{P_{ni}} = \frac{e^{\alpha (C_{mi}+1)}}{e^{\alpha C_{ni}}} = e^{\alpha C_{mi} + \alpha - \alpha C_{ni}}$$

Now increase cost of m by $1.

$$\boxed{OR_{m|n} = \frac{\Omega'_{m|n}}{\Omega_{m|n}} = \frac{e^{\alpha C_{mi} + \alpha - \alpha C_{ni}}}{e^{\alpha C_{mi} - \alpha C_{ni}}} = e^{\alpha}}$$

Suppose both increased by one dollar?

Logit w/ **Both** Individual and Choice Characteristics

$$U_{1i} = \mathbf{x}_i\boldsymbol{\beta}_1 + \mathbf{z}_{1i}\boldsymbol{\alpha} + \varepsilon_{1i}$$
$$U_{2i} = \mathbf{x}_i\boldsymbol{\beta}_2 + \mathbf{z}_{2i}\boldsymbol{\alpha} + \varepsilon_{2i}$$

The utility functions for the choices have two types of variables. Note that $\mathbf{z}_{ji}$ has a double subscript, because choice characteristics vary by *choice* (j) and by *person* (i). $\boldsymbol{\alpha}$ has no subscript. Why?

$$\begin{aligned}
\Pr\left(Y_i = 1 \mid \mathbf{x}_i, \mathbf{z}_{ji}\right) &= \Pr\left(U_{1i} > U_{2i}\right) \\
&= \Pr\left[\left(\mathbf{x}_i\boldsymbol{\beta}_1 + \mathbf{z}_{1i}\boldsymbol{\alpha} + \varepsilon_{1i}\right) > \left(\mathbf{x}_i\boldsymbol{\beta}_2 + \mathbf{z}_{2i}\boldsymbol{\alpha} + \varepsilon_{2i}\right)\right] \\
&= \Pr\left[\left(\varepsilon_{1i} - \varepsilon_{2i}\right) > \left(\mathbf{x}_i\left(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1\right) + \boldsymbol{\alpha}\left(\mathbf{z}_{2i} - \mathbf{z}_{1i}\right)\right)\right]
\end{aligned}$$

Are the betas identified? Are the alphas identified?

$$P_{1i} = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}_1 + \mathbf{z}_{1i}\boldsymbol{\alpha}}}{e^{\mathbf{x}_i\boldsymbol{\beta}_1 + \mathbf{z}_{1i}\boldsymbol{\alpha}} + e^{\mathbf{x}_i\boldsymbol{\beta}_2 + \mathbf{z}_{2i}\boldsymbol{\alpha}}} \qquad P_{2i} = \frac{e^{\mathbf{x}_i\boldsymbol{\beta}_2 + \mathbf{z}_{2i}\boldsymbol{\alpha}}}{e^{\mathbf{x}_i\boldsymbol{\beta}_1 + \mathbf{z}_{1i}\boldsymbol{\alpha}} + e^{\mathbf{x}_i\boldsymbol{\beta}_2 + \mathbf{z}_{2i}\boldsymbol{\alpha}}}$$

Generalizing

$$U_{1i} = \mathbf{x}_i \boldsymbol{\beta}_1 + \mathbf{z}_{1i}\boldsymbol{\alpha} + \varepsilon_{1i}$$

$$U_{2i} = \mathbf{x}_i \boldsymbol{\beta}_2 + \mathbf{z}_{2i}\boldsymbol{\alpha} + \varepsilon_{2i}$$

$$\vdots$$

$$U_{Ji} = \mathbf{x}_i \boldsymbol{\beta}_J + \mathbf{z}_{Ji}\boldsymbol{\alpha} + \varepsilon_{Ji}$$

Assuming IID Type I extreme distributions for the ε's leads to the following form for the probability of outcome 1, and the other are analogous.

Identifying assumption:

$$\boldsymbol{\beta}_1 = \mathbf{0}$$

$$P_{ji} = \frac{e^{\mathbf{x}_i \boldsymbol{\beta}_j + \mathbf{z}_{ji}\boldsymbol{\alpha}}}{\sum_{j=1}^{J} e^{\mathbf{x}_i \boldsymbol{\beta}_j + \mathbf{z}_{ji}\boldsymbol{\alpha}}}$$

From here, construct the likelihood function and use MLE to get the estimates of $\boldsymbol{\beta}_j$ and $\boldsymbol{\alpha}$.

Allowing Different Intercepts

. asclogit picked price, case(person) alternatives(rectype)

```
Alternative-specific conditional logit         Number of obs      =      3,000
Case variable: person                          Number of cases    =       1000

Alternative variable: rectype                  Alts per case: min =          3
                                                              avg =        3.0
                                                              max =          3


                                                  Wald chi2(1)    =       0.65
Log likelihood = -1094.2186                       Prob > chi2     =     0.4211

------------------------------------------------------------------------------
      picked |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rectype      |
       price |  -.0476118   .0591766    -0.80   0.421    -.1635958    .0683722
-------------+----------------------------------------------------------------
1            |  (base alternative)
-------------+----------------------------------------------------------------
2            |
       _cons |  -.2788324   .3065368    -0.91   0.363    -.8796335    .3219686
-------------+----------------------------------------------------------------
3            |
       _cons |  -.5446558   .4203867    -1.30   0.195    -1.368599    .2792869
------------------------------------------------------------------------------
```

Predictions for Person 1

$$\hat{U}_{1i} = \hat{\alpha}C_{1i} = 0 - 0.0476(7.05) = -0.336$$

$$\hat{U}_{2i} = \hat{\alpha}C_{1i} = -0.279 - 0.0476(1.40) = -0.345$$

$$\hat{U}_{3i} = \hat{\alpha}C_{1i} = -0.545 - 0.0476(0) = -0.545$$

$$\hat{P}_1 = \frac{e^{\hat{U}_1}}{e^{\hat{U}_1} + e^{\hat{U}_2} + e^{\hat{U}_3}}$$

Etc., etc. Verify the results below. Note: if price is the same, the probabilities will be different because each alternative has a different intercept.

```
. predict prob

. list choice price  picked prob if person==1

     +--------------------------------------+
     | choice       price   picked       prob |
     |--------------------------------------|
  1. |   Walk   7.050649        0   .3569322 |
  2. |   Walk   1.400565        0   .3534437 |
  3. |   Walk          0        1   .2896242 |
     +--------------------------------------+
```

Now add in individual characteristics

. asclogit picked price, case(person) alternatives(rectype) casevars(income)

```
Alternative-specific conditional logit         Number of obs      =      3000
Case variable: person                          Number of cases    =      1000
Alternative variable: rectype                  Alts per case: min =         3
                                                              avg =       3.0
                                                              max =         3
                                                  Wald chi2(3)    =    282.79
Log likelihood = -829.89605                       Prob > chi2     =    0.0000
------------------------------------------------------------------------------
      picked |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rectype      |
       price |  -.0883262   .0712086    -1.24   0.215    -.2278925    .0512401
-------------+----------------------------------------------------------------
rectype2     |
      income |  -.1870799   .0139695   -13.39   0.000    -.2144597   -.1597001
       _cons |   3.865813   .4839139     7.99   0.000     2.917359    4.814267
-------------+----------------------------------------------------------------
rectype3     |
      income |  -.2799832   .0168107   -16.66   0.000    -.3129316   -.2470349
       _cons |   4.978513   .6055072     8.22   0.000     3.791741    6.165286
------------------------------------------------------------------------------
```

Predicted Utilities for *income* = \$20, P_{car}= \$7.00, P_{bus}= \$2.00, P_{walk}= \$0

$$\hat{U}_{1i} = \mathbf{x}_i \hat{\boldsymbol{\beta}}_1 + \mathbf{z}_{1i} \hat{\boldsymbol{\alpha}}$$

$$\hat{U}_{2i} = \mathbf{x}_i \hat{\boldsymbol{\beta}}_2 + \mathbf{z}_{2i} \hat{\boldsymbol{\alpha}}$$

$$\hat{U}_{3i} = \mathbf{x}_i \hat{\boldsymbol{\beta}}_3 + \mathbf{z}_{3i} \hat{\boldsymbol{\alpha}}$$

Probabilities

$$\hat{P}_1 = \frac{e^{-0.616}}{e^{-0.616} + e^{-0.050} + e^{-0.621}} = \frac{0.540}{2.029} = 0.266$$

$$\hat{P}_2 = \frac{e^{-0.050}}{e^{-0.616} + e^{-0.050} + e^{-0.621}} = \frac{0.951}{2.029} = 0.469$$

$$\hat{P}_3 = \frac{e^{-0.621}}{e^{-0.616} + e^{-0.050} + e^{-0.621}} = \frac{0.537}{2.029} = 0.265$$

$$\sum_{j=1}^{J} \hat{P}_{ji} = 1$$

Summary

- Binomial logit can be derived from a random utility framework.
- Binomial logit is a special case of multinomial logit
- Multinomial logit is based on individual characteristics
- Conditional logit conditions on choice characteristics (or both individual and choice characteristics)
- All these models implicitly assume IIA (see Alvarez and Nagler)