Alternative Specific Logit

The example for this section comes from Cameron and Trivedi's excellent book, *Microeconometrics Using Stata*. The data are originally from Herriges and Kling (REStat 1999) and is available from the Stata website http://www.stata-press.com/data/musr.html.

There are 1182 observations and there is one observation for each individual. The choice being made is the mode of fishing; one can fish from the beach, a pier, private boat, or charter boat. There are choice specific variables included as well: the catch rate for each choice and the price for each choice. Income of each individual is also included.

First, describe the data:

describe

```
. describe

Contains data from C:\Users\Lee\Documents\mus15data.dta
  obs:         1,182
 vars:            16                          12 May 2008 20:46
 size:        80,376 (99.8% 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 $
```

You can see that the data set contains the choice (`mode`) and its `price` and `catch` rate. It also breaks the mode down into a set of indicator variables (`dbeach`, `dpier`, `dprivate`, `dcharter`). It includes the prices for each of the other alternatives and their respective catch rates. Finally is the individual specific variable, `income`.

To verify this, list the first few observations:

list in 1/3

```
. list in 1/3
```

	mode	price	crate	dbeach	dpier	dprivate	dcharter	pbeach	ppier
1.	charter	182.93	.5391	0	0	0	1	157.93	157.93
	pprivate	**pcharter**	**qbeach**	**qpier**	**qprivate**	**qcharter**	**income**		
	157.93	182.93	.0678	.0503	.2601	.5391	7.083332		
2.	charter	34.534	.4671	0	0	0	1	15.114	15.114
	10.534	34.534	.1049	.0451	.1574	.4671	1.25		
3.	private	24.334	.2413	0	0	1	0	161.874	161.874
	24.334	59.334	.5333	.4522	.2413	1.0266	3.75		

The first individual fishes by charter at a price of $182.93 and catch rate of 0.5391. The prices of the alternatives and catch rates are given.

You can clean this up a bit using the clean option, here done on a subset of the variables

```
. list mode price crate pbeach ppier pprivate pcharter in 1/3, clean

        mode    price    crate    pbeach     ppier   pprivate   pcharter
  1.  charter   182.93    .5391    157.93    157.93     157.93     182.93
  2.  charter   34.534    .4671    15.114    15.114     10.534     34.534
  3.  private   24.334    .2413   161.874   161.874     24.334     59.334
```

The **tabulate** command gives you an idea about the relative frequencies of each choice.

tabulate mode

```
. tabulate mode

    Fishing |
       mode |      Freq.     Percent        Cum.
------------+-----------------------------------
      beach |        134       11.34       11.34
       pier |        178       15.06       26.40
    private |        418       35.36       61.76
    charter |        452       38.24      100.00
------------+-----------------------------------
      Total |      1,182      100.00
```

Most folks fish by charter (38%) or private boat (35%).

Another summary can be created using the table command.

```
. table mode, contents(N income mean income sd income)

-------------------------------------------------
Fishing   |
mode      |    N(income)  mean(income)    sd(income)
----------+--------------------------------------
    beach |          134      4.051617       2.50542
     pier |          178      3.387172      2.340324
  private |          418      4.654107      2.777898
  charter |          452      3.880899      2.050029
-------------------------------------------------
```

You can also use table to summarize the alternative specific variables:

```
. table mode, contents(mean pbeach mean ppier mean pprivate mean pcharter)
format(%6.0f)

-----------------------------------------------------------------------
Fishing   |
mode      |   mean(pbeach)     mean(ppier)  mean(pprivate)  mean(pcharter)
----------+------------------------------------------------------------
    beach |             36              36              98             125
     pier |             31              31              82             110
  private |            138             138              42              71
  charter |            121             121              45              75
-----------------------------------------------------------------------
```

It is fairly clear that the beach and pier carry the same price. Also, it looks like people are fairly price sensitive. Private boaters and charters avoid the beach and pier when they are expensive. Beach and pier fishermen face higher prices for boat fishing.

```
. table mode, contents(mean qbeach mean qpier mean qprivate mean qcharter)
format(%6.4f)

-----------------------------------------------------------------------
Fishing   |
mode      |   mean(qbeach)     mean(qpier)  mean(qprivate)  mean(qcharter)
----------+------------------------------------------------------------
    beach |         0.2792          0.2190          0.1594          0.5176
     pier |         0.2614          0.2025          0.1501          0.4981
  private |         0.2083          0.1298          0.1775          0.6539
  charter |         0.2519          0.1595          0.1772          0.6915
-----------------------------------------------------------------------
```

Multinomial Logit

First, we'll estimate a MNL model. The choice depends on individual characteristics, which in this case is income.

```
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]
-------------+----------------------------------------------------------------
beach        |  (base outcome)
-------------+----------------------------------------------------------------
pier         |
      income |  -.1434029   .0532884    -2.69   0.007    -.2478463   -.0389595
       _cons |   .8141503    .228632     3.56   0.000     .3660399    1.262261
-------------+----------------------------------------------------------------
private      |
      income |   .0919064   .0406637     2.26   0.024     .0122069    .1716058
       _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
------------------------------------------------------------------------------
```

Now, in terms of relative risk ratios

```
. mlogit, rrr
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 |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
beach        |  (base outcome)
-------------+----------------------------------------------------------------
pier         |
      income |   .8664049   .0461693    -2.69   0.007     .7804799    .9617896
-------------+----------------------------------------------------------------
private      |
      income |   1.096262   .0445781     2.26   0.024     1.012282     1.18721
-------------+----------------------------------------------------------------
charter      |
      income |   .9688554    .040543    -0.76   0.450     .8925639    1.051668
------------------------------------------------------------------------------
```

So, a $1000 increase in income increases the odds of private fishing rather than the beach by a small amount and reduces the probability of fishing the pier or by charter relative to fishing the beach.

You can get similar information using marginal effects. For instance, the average marginal effect of a $1000 increase in income on the probability of choosing charter:

```
. margins, dydx(income) predict(outcome(charter))

Average marginal effects                          Number of obs   =       1182
Model VCE    : OIM

Expression   : Pr(mode==charter), predict(outcome(charter))
dy/dx w.r.t. : income

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |  -.0111519   .0059441    -1.88   0.061     -.022802    .0004983
------------------------------------------------------------------------------
```

As expected, it is negative, though zero is within the 95% confidence interval.

Conditional logit

This model is probably better described as McFadden's alternative specific conditional logit. It is a special case of conditional logit where there are no characteristics (individual specific variables) and only attributes (choice specific variables). The more general conditional logit combines the two. The Stata command used to estimate this model is **asclogit** (although you can use **clogit** if you create interactions with the indicators).

First, you have to get the data into long-form. In this form, each observation contains the data for each alternative. With 4 choices, there will be 4 observations for each individual. The first thing to do is to create an id number for each individual. Then use the **reshape long** command as shown below.

```
generate id = _n
reshape long d p q, i(id) j(fishmode beach pier private charter) string
list in 1/4
```

The syntax deserves some explanation. [Stata Data Manual, pp. 481-482] Before using **reshape**, you need to determine whether the data are in long or wide form. You also must determine the logical observation (**i**) and the subobservation (**j**) by which to organize the data. Suppose that you had the following data, which could be organized in wide or long form as follows:

i		..... X_{ij}		
id	sex	inc80	inc81	inc82
1	0	5000	5500	6000
2	1	2000	2200	3300
3	0	3000	2000	1000

i	j		X_{ij}
id	year	sex	inc
1	80	0	5000
1	81	0	5500
1	82	0	6000
2	80	1	2000
2	81	1	2200
2	82	1	3300
3	80	0	3000
3	81	0	2000
3	82	0	1000

Given these data, you could use `reshape` to convert from one form to the other:

```
. reshape long inc, i(id) j(year)          /* goes from left form to right */
. reshape wide inc, i(id) j(year)          /* goes from right form to left */
```

So for our purposes, we have generated the **id** variable that is to be used as what Stata refers to as the "logical" observation, which is identified by the 'option' **i(id)**. The subovservations consist of the indicator, price, and catch-rate data that are prefixed by and renamed to **d**, **p**, and **q**. The **j()** command includes the variable names, or in this case, the fragment of the name, for the variables to create. The string option is given because the **fishmode** variable is a string. Here is what the reshaping does to the data:

```
. list in 1/8, table sepby(id)

     +-------------------------------------------------------------------------+
     | id   fishmode      mode    price   crate   d        p       q     income |
     |-------------------------------------------------------------------------|
  1. |  1      beach   charter   182.93   .5391   0   157.93   .0678   7.083332 |
  2. |  1    charter   charter   182.93   .5391   1   182.93   .5391   7.083332 |
  3. |  1       pier   charter   182.93   .5391   0   157.93   .0503   7.083332 |
  4. |  1    private   charter   182.93   .5391   0   157.93   .2601   7.083332 |
     |-------------------------------------------------------------------------|
  5. |  2      beach   charter   34.534   .4671   0   15.114   .1049       1.25 |
  6. |  2    charter   charter   34.534   .4671   1   34.534   .4671       1.25 |
  7. |  2       pier   charter   34.534   .4671   0   15.114   .0451       1.25 |
  8. |  2    private   charter   34.534   .4671   0   10.534   .1574       1.25 |
     +-------------------------------------------------------------------------+
```

The variables **mode**, **price** and **crate** refer to the choice that is made by the individual. The variables **d**, **p**, and **q** refer to the alternatives faced by the individual; these are the ones that will be used for **asclogit**.

McFadden's model only includes variables that are choice specific.

```
. * Conditional logit with alternative specific variables
. asclogit d p q, case(id) alternatives(fishmode)

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(2)       =     229.35
Log likelihood = -1230.7838                       Prob > chi2        =     0.0000
------------------------------------------------------------------------------
           d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
fishmode     |
           p |  -.0247896   .0017044   -14.54   0.000    -.0281301    -.021449
           q |   .3771689   .1099707     3.43   0.001     .1616303    .5927074
-------------+----------------------------------------------------------------
beach        |   (base alternative)
-------------+----------------------------------------------------------------
charter      |
       _cons |   1.498888   .1329328    11.28   0.000     1.238345    1.759432
-------------+----------------------------------------------------------------
pier         |
       _cons |   .3070552   .1145738     2.68   0.007     .0824947    .5316158
-------------+----------------------------------------------------------------
private      |
       _cons |   .8713749   .1140428     7.64   0.000     .6478551    1.094895
------------------------------------------------------------------------------
```

Notice that the 'slopes' are common to price (**p**) and catch-rate (**q**). Each choice has a separate constant. The conditional logit model does not use the **margins** command, instead it has its own post-estimation command **estat mfx**. Here is the marginal effect of an increase in price evaluated at the mean prices.

```
. estat mfx, varlist(p)

Pr(choice = beach|1 selected) =  .0546071
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.     z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
p            |
       beach |  -.00128    .00012  -10.66   0.000  -.001515  -.001044    103.42
     charter |  .000614    .00006   10.25   0.000   .000497   .000732    84.379
        pier |  .000098   .000017    5.88   0.000   .000065    .00013    103.42
     private |  .000568   .000056   10.16   0.000   .000458   .000677    55.257
------------------------------------------------------------------------------

Pr(choice = charter|1 selected) = .45376978
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.     z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
p            |
```

```
       beach |  .000614    .00006   10.25   0.000   .000497   .000732    103.42
     charter | -.006144   .000435  -14.12   0.000  -.006997  -.005291    84.379
        pier |  .000811   .000071   11.42   0.000   .000671    .00095    103.42
     private |   .00472   .000437   10.80   0.000   .003863   .005576    55.257
-------------------------------------------------------------------------------

Pr(choice = pier|1 selected) = .07206028
-------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.     z     P>|z|  [    95% C.I.    ]       X
-------------+-----------------------------------------------------------------
p            |
       beach |  .000098   .000017    5.88   0.000   .000065    .00013    103.42
     charter |  .000811   .000071   11.42   0.000   .000671    .00095    84.379
        pier | -.001658   .000137  -12.07   0.000  -.001927  -.001389    103.42
     private |  .000749   .000066   11.30   0.000   .000619   .000879    55.257
-------------------------------------------------------------------------------

Pr(choice = private|1 selected) = .41956284
-------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.     z     P>|z|  [    95% C.I.    ]       X
-------------+-----------------------------------------------------------------
p            |
       beach |  .000568   .000056   10.16   0.000   .000458   .000677    103.42
     charter |   .00472   .000437   10.80   0.000   .003863   .005576    84.379
        pier |  .000749   .000066   11.30   0.000   .000619   .000879    103.42
     private | -.006037   .000437  -13.82   0.000  -.006893  -.005181    55.257
-------------------------------------------------------------------------------
```

Notice that for those that choose beach, and increase in the beach price reduces the probability of choosing beach and this is distributed across the other choices.

Adding characteristics (case specific variables)

The conditional logit can also be estimated with individual specific variables like **income**. Stata calls the characteristics "cases," and to use them they need to be included in the **casevars( )** option.

```
. asclogit d p q, case(id) alternatives(fishmode) casevars(income)

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
------------------------------------------------------------------------------
```

Income and constant have separate coefficients for each choice while the attributes have common slopes.

Below I've computed marginal effects (again, at the means) for the income variable:

```
. estat mfx, varlist(income)

Pr(choice = beach|1 selected) = .05248806
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
casevars     |
      income | -.000721   .002319   -0.31   0.756  -.005266   .003823    4.0993
------------------------------------------------------------------------------

Pr(choice = charter|1 selected) = .46206853
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
casevars     |
      income | -.021734    .00666   -3.26   0.001  -.034787   -.00868    4.0993
------------------------------------------------------------------------------

Pr(choice = pier|1 selected) = .06584968
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
casevars     |
      income | -.009306   .002719   -3.42   0.001  -.014635  -.003977    4.0993
```

```
------------------------------------------------------------------------------

Pr(choice = private|1 selected) = .41959373
------------------------------------------------------------------------------
variable     |   dp/dx   Std. Err.    z     P>|z|  [    95% C.I.    ]       X
-------------+----------------------------------------------------------------
casevars     |
      income |  .031761   .006554    4.85   0.000   .018915   .044608    4.0993
------------------------------------------------------------------------------
```

In this example, only the probability of fishing from a private boat increases with income.

Finally, you can also use **asclogit** to estimate multinomial logit. This is handy if the variables are in long-form and you don't want to reshape them.

```
. * MNL is CL with no alternative specific regressors
. asclogit d, case(id) alternatives(fishmode) casevars(income) basealternative(beach)

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(3)       =      37.70
Log likelihood = -1477.1506                    Prob > chi2        =     0.0000

------------------------------------------------------------------------------
           d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
beach        |   (base alternative)
-------------+----------------------------------------------------------------
charter      |
      income |    -.03164   .0418463    -0.76   0.450    -.1136572    .0503773
       _cons |   1.341292   .1945167     6.90   0.000     .9600459    1.722537
-------------+----------------------------------------------------------------
pier         |
      income |  -.1434028   .0532884    -2.69   0.007    -.2478462   -.0389595
       _cons |     .81415   .2286319     3.56   0.000     .3660396     1.26226
-------------+----------------------------------------------------------------
private      |
      income |   .0919063   .0406637     2.26   0.024     .0122068    .1716057
       _cons |    .738921   .1967309     3.76   0.000     .3533355    1.124507
------------------------------------------------------------------------------
```

These results match those from MNL.

do-file

```
use C:\Users\Lee\Documents\mus15data.dta, clear

describe

list mode price crate pbeach ppier pprivate pcharter in 1/4, clean

* Multinomial Logit -- case specific vars only in short form
mlogit mode income, baseoutcome(1)
mlogit, rrr
margins, dydx(income) predict(outcome(charter))

* Convert data to long form for asclogit
generate id = _n
reshape long d p q, i(id) j(fishmode beach pier private charter) string
list in 1/8

* Drop the variables that don't make sense in long form
drop mode price crate

* Conditional logit with alternative specific variables
asclogit d p q, case(id) alternatives(fishmode)
estat mfx, varlist(p)

* Conditional logit with case and alternative specific variables
asclogit d p q, case(id) alternatives(fishmode) casevars(income)

* Predicted probabilities of choice of each mode and compare to actual freqs
predict pasclogit, pr
table fishmode, contents(mean d mean pasclogit sd pasclogit) cellwidth(15)

* Marginal effect at mean of change in price
estat mfx, varlist(p)
estat mfx, varlist(income)

* MNL is CL with no alternative specific regressors
asclogit d, case(id) alternatives(fishmode) casevars(income) basealternative(beach)
```