POLS570 - Limited Dependent Variables
October 30, 2006
Conditional Logit

1   Introduction

You should understand one key thing about the conditional logit model:

It is exactly the same as the multinomial logit model. Period.

This means that:

- The formulas are very similar-looking, and can be shown to be mathematically equivalent (e.g. Long, Maddala),

- Estimation is done exactly the same way,

- You will get identical results from MNL and CL if you run the same model,

- All of the ways in which you can interpret a MNL model, you can also interpret a conditional logit.

That said, it turns out that the CL is actually more flexible than the MNL model in most circumstances. So, why the differences?

2   Conditional Logit

The MNL model measures the impact of observation-specific variables on the probability of choosing one of J discrete outcomes. In the MNL model, the data look like this:

```
OBS     CHOICE      X1          X2
---------------------------------
1        0          47          0
2        2          13          1
3        1          82          1
.        .          .           .
.        .          .           .
.        .          .           .
```

etc. This is familiar: there is one line of data per observation, and covariates (that is Xs) vary across individuals/cases, but not across choices. Estimation yields separate effects for each variable on each choice.

But what if we have some variable (say, Z_{ij}) that varies across *choices*, or across *both* observations and choices?

- E.g. In a study of Presidential elections, we might want to incorporate some measure of the voter's evaluation of each candidate, which will vary across choices/candidates.

- Or: Assigning opinions to Supreme Court justices, on the basis of how close each is to the opinion assigner...

This gives us what is called the *conditional logit* ("CL") model...

CL allows us to estimate the effect of choice-specific variables on the probability of choosing a particular alternative. To do this, we need to restructure the data into what we might call "long" format...

OBS	Alternative	CHOICE	X1	X2	Z
1	0	1	47	0	10
1	1	0	47	0	24
1	2	0	47	0	49
2	0	0	13	1	47
2	1	0	13	1	51
2	2	1	13	1	8
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

etc. Here:

- OBS indexes the observation (i.e., case).

- Alternative indexes the (possible) choice: 0, 1 or 2.

- CHOICE is a dummy variable, equaling 1 if that observation chose that alternative, and 0 otherwise.

- X1 and X2 don't change across alternatives.

- Z *does* change across alternatives...

2

The conditional logit estimates a model with the basic probability function:

$$Pr(Y_{ij} = j) = \frac{exp(Z_{ij}\gamma)}{\sum_{j=1}^{J} exp(Z_{ij}\gamma)} \tag{1}$$

where Z_{ij} is a vector of independent variables that varies across both observations i and choice alternatives j, and γ is a $k \times 1$ vector of coefficients to be estimated.

Note a couple things:

- Because the Zs vary across both observations and alternatives, we assume that the effect of a change in Z on the probability of choosing a particular alternative is constant across alternatives.

- Hence, we only estimate k parameters (not $(J-1)k$, as in the MNL model).

So, the conditional logit model is useful when we want to estimate the impact of alternative- or choice-specific variables on a polychotomous dependent variable. E.g.:

- Choices of transportation alternatives, according to how long each alternative takes, or

- Choosing political candidates on the basis of where each respondent places that candidate ideologically, relative to him or her; or "feeling thermometers"...

3 Modeling Observation- and Choice-Specific Effects

Question: What happens if we want to estimate joint effects of variables which are constant across choices (e.g. demographic effects) and those which vary across alternatives (e.g. candidate distance)?

- Suppose we think, say, that both party identification and "feelings" affect candidate choice, or

- that a Congressperson's decision to run for reelection, seek higher office, or retire depends on both individual factors (e.g. his/her age) and the expected likelihood of winning in each case (which varies by alternative as well as by candidate).

The simplest kind of alternative-specific effect is simply a *fixed effect*...

- The CL model in the form given above doesn't include a constant term...

- We might (probably would) expect some alternatives to be more or less likely than others, just in general...

- SO we can include indicator variables for each possible alternative (a la "fixed effects" for each alternative).

In fact, we can then use these fixed effects to "trick" the CL into estimating a model with independent variables which do not vary across alternatives. To understand why this is, think about the notion of an interaction term with a dummy variable...

- Term allows the effect of some variable to have a different effect on the dependent variable for each of two groups.

- E.g. ideology might have a different effect on judicial voting for activist judges than it does for restraintists.

- Here, we can use this idea:

    - The interaction of an observation-specific variable with an alternative-specific constant allows the observation-specific variable's effect to be different for that alternative than for the others.
    - This is how we can get the CL model to estimate effects for observation-specific variables

It turns out that, if we do this, its possible (in fact, easy) to get results identical to those you'd get from running a MNL model...

4   MNL and CL: An Example

This example illustrates how to estimate conditional logit models in Stata, including models with either or both alternative-specific and observation-specific independent variables. It also illustrates the mathematical equivalence of the MNL and CL specifications. The data are completely made-up, and take on the "long" form, with one observation for each case for each alternative on the dependent variable. There are 1000 observations, and three

possible "choices" for each observation, yielding 3000 lines of data. The variables consist of Y, a dummy variable coded 1 if that observation "chose" that alternative and 0 otherwise, and two independent variables. Also included is the variable choice, coded $\{0, 1, 2\}$ which indicates which of the three outcomes was chosen (this variable thus does not vary across choices). There are 377 "0s", 308 "1s", and 315 "2s" in these data. The independent variable X also does not vary across alternatives, while the independent variable Z varies across both observations and alternatives. In an election context, we might think of X as some individual (voter) – specific characteristic, such as race, age, gender, income, etc., while Z represents some attribute specific to both the voter and the candidate (e.g. the perceived ideological distance between voter and candidate). Two other variables serve as identifiers: id is a unique identifier for each observation, while alt is a three-category $\{0, 1, 2\}$ nominal variable indexing the three possible alternatives. By way of example, the first three "observations" in the data contain the following:

```
. list

      id    alt    choice    Y      X          Z
1.  1     0      0       1    1.14518   -.4012452
2.  1     1      0       0    1.14518    1.870998
3.  1     2      0       0    1.14518   -1.104332
4.  2     0      2       0    3.422003   1.508909
5.  2     1      2       0    3.422003   1.968283
6.  2     2      2       1    3.422003    .6115509
7.  3     0      2       0    6.435474   1.606365
8.  3     1      2       0    6.435474   1.693768
9.  3     2      2       1    6.435474   2.195817
```

So, for observation 1, choice $\{0\}$ was chosen; observation 2 chose alternative $\{2\}$, etc. Lets begin by estimating a multinomial logit model of the choice decision, as a function of the independent variable X. As is common in other software programs, Stata's -mlogit- command doesn't allow inclusion of choice-specific independent variables, so we cannot assess the impact of Z in this manner. Since each "observation" has three lines of data associated with it, and since neither choice nor X vary across observations, one can select out any 1000 lines of data, one for each observation, to do this analysis. Here I do this by restricting the analysis to those data points where alt = 0:

```
. mlogit choice X if alt==0
```

```
Iteration 0: Log Likelihood =-1094.3677...
Iteration 6: Log Likelihood = -477.8912

Multinomial regression                    Number of obs   =        1000
                                          LR chi2(2)      =     1232.95
                                          Prob > chi2     =      0.0000
Log likelihood =  -477.8912               Pseudo R2       =      0.5633
------------------------------------------------------------------------
   choice |    Coef.   Std. Err       z   P>|z|    [95% Conf. Interval]
----------+-------------------------------------------------------------
1         |
        X |   1.96101  .1441057   13.608  0.000    1.678568    2.243452
    _cons |-4.157174  .3165533  -13.133  0.000   -4.777607   -3.536741
----------+-------------------------------------------------------------
2         |
        X | 3.909589  .2131314   18.344  0.000    3.491859    4.327319
    _cons |-11.96931  .7020588  -17.049  0.000   -13.34532   -10.5933
------------------------------------------------------------------------
(Outcome choice==0 is the comparison group)
```

X has a positive, significant effects on the probabilities of choosing alternatives {1} or {2} relative to alternative {0}. If we want to know the effect of Z on choosing each of the various alternatives, we use Stata's -clogit- command:

```
. clogit Y Z, gr(id)

Iteration 0:   log likelihood = -1098.3347...
Iteration 2:   log likelihood = -1097.4566

Conditional (fixed-effects)               Number of obs   =        3000
    logistic regression                   LR chi2(1)      =        2.31
                                          Prob > chi2     =      0.1284
Log likelihood = -1097.4566               Pseudo R2       =      0.0011
------------------------------------------------------------------------
     Y |    Coef.   Std. Err.    z    P>|z|      [95% Conf. Interval]
-------+----------------------------------------------------------------
     Z | .0593299  .0390661   1.519  0.129     -.0172382     .135898
------------------------------------------------------------------------
```

Note the use of the `-gr()-` (for "group") option to indicate the variable which uniquely identifies each observation in the data. Here, Z appears to have a small positive impact on the choice of alternatives; as Z increases for a particular choice, the probability of its being chosen also increases slightly. Note also the lack of a constant in this model: if we believe that individual choices are, in general, more or less likely to be chosen than others, we can capture this effect by including *choice-specific constants* in the CL model:

```
. gen ch1=0
. replace ch1=1 if alt==1
(1000 real changes made)

. gen ch2=0
. replace ch2=1 if alt==2
(1000 real changes made)

. clogit Y Z ch1 ch2, gr(id)
Iteration 0: Log Likelihood =-1095.9458...
Iteration 2: log likelihood = -1093.2573

Conditional (fixed-effects)                 Number of obs   =        3000
   logistic regression                      LR chi2(3)      =       10.71
                                            Prob > chi2     =      0.0134
Log likelihood = -1093.2573                 Pseudo R2       =      0.0049
------------------------------------------------------------------------
    y |     Coef.  Std. Err.     z    P>|z|     [95% Conf. Interval]
------+-----------------------------------------------------------------
    Z |  .0582532  .0391291   1.489  0.137    -.0184384     .1349447
  ch1 | -.2016517  .0768495  -2.624  0.009    -.3522739    -.0510294
  ch2 | -.1781627  .0763861  -2.332  0.020    -.3278768    -.0284487
------------------------------------------------------------------------
```

Both alternative {1} and {2} are less likely to be chosen, on average, as we would expect given that they are less common outcomes than is alternative {0}. While this is also informative, we may be especially interested in the effect of, say Z on Y after holding the X variable(s) constant. To do this, we need to include both X and Z in the same model; however, we cannot do this with the data in their current form, since the values of X do not vary across alternatives:

```
. clogit Y Z X ch1 ch2, gr(id)
```

```
Note: X omitted due to no within-group variance.

Iteration 0: Log Likelihood =-1095.9458...
Iteration 2: log likelihood = -1093.2573

Conditional (fixed-effects)                  Number of obs   =       3000
   logistic regression                       LR chi2(3)      =      10.71
                                             Prob > chi2     =     0.0134
Log likelihood = -1093.2573                  Pseudo R2       =     0.0049
------------------------------------------------------------------------
     y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------+-----------------------------------------------------------------
     Z |   .0582532   .0391291    1.489   0.137     -.0184384    .1349447
   ch1 |  -.2016517   .0768495   -2.624   0.009     -.3522739   -.0510294
   ch2 |  -.1781627   .0763861   -2.332   0.020     -.3278768   -.0284487
------------------------------------------------------------------------
```

The key to getting around these problems is to recognize that variables
which do not vary across alternatives must be allowed to exert a separate
impact on each alternative (a la MNL). Including the "fixed effects" for
the alternatives turns out to be the key to doing this. This can be done
by simply interacting each of the observation-specific variables with each of
the alternative-specific constants, then including both direct and interactive
effects in the conditional logit model:

```
. gen Xalt1=X*ch1
. gen Xalt2=X*ch2
. clogit Y Z Xalt1 Xalt2 ch1 ch2, gr(id)

Iteration 0: Log Likelihood =-885.08548...
Iteration 6:   log likelihood = -476.47227

Conditional (fixed-effects)                  Number of obs   =       3000
   logistic regression                       LR chi2(5)      =    1244.28
                                             Prob > chi2     =     0.0000
Log likelihood = -476.47227                  Pseudo R2       =     0.5663
------------------------------------------------------------------------
     Y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------+-----------------------------------------------------------------
```

8

```
     Z |  .1010649    .0601396    1.681   0.093    -.0168066    .2189364
   ch1 | -4.16007     .3174664   -13.104   0.000    -4.782293   -3.537848
   ch2 |-12.02097     .7064777   -17.015   0.000   -13.40564    -10.6363
 Xalt1 | 1.964062     .1444774    13.594   0.000     1.680891    2.247232
 Xalt2 | 3.923085     .2142272    18.313   0.000     3.503207    4.342962
-----------------------------------------------------------------------
```

Alternative {0} is now the omitted, baseline category, and the coefficients
on the two X interactions indicate the effect of the X variable on choosing
each of the two other alternatives {1} and {2}. By interacting X with the
two alternative-specific constants, we have allowed the effect of X to vary
across alternatives. Note also that the effect of Z appears to be somewhat
greater, controlling for X.

Finally, watch what happens when we omit Z from this last analysis:

```
. clogit y Xalt1 Xalt2 ch1 ch2, gr(id)
Iteration 0: Log Likelihood =-989.89452...
Iteration 6:   log likelihood =  -477.8912

Conditional (fixed-effects)                 Number of obs   =        3000
    logistic regression                     LR chi2(4)      =     1241.44
                                            Prob > chi2     =      0.0000
Log likelihood =  -477.8912                 Pseudo R2       =      0.5650
-----------------------------------------------------------------------
     Y |     Coef.  Std. Err.      z     P>|z|    [95% Conf. Interval]
-------+---------------------------------------------------------------
   ch1 | -4.157174   .3165545    -13.133   0.000    -4.77761    -3.536739
   ch2 | -11.96931   .7020656    -17.049   0.000   -13.34533   -10.59328
 Xalt1 |   1.96101   .1441062     13.608   0.000     1.678567    2.243453
 Xalt2 | 3.909589    .2131331     18.343   0.000     3.491856    4.327322
-----------------------------------------------------------------------
```

The log-likelihood, coefficient estimates, standard errors, significance lev-
els, and everything else are exactly the same as those obtained using the
-mlogit- command, above. The alternative-specific constants serve as sep-
arate "constant terms" for each possible choice, while the interactive terms
indicate the (relative) effects of X on each. This illustrates, in a very graphic
fashion, that these are in fact the same model, and will yield identical results
if specified in the same way.

5   Conditional Logit: Another Example

Maltzmann and Wahlbeck study opinion assignment in the Rehnquist Court.[1] Their data are majority opinion assignments by Chief Justice Rehnquist ($N = 316$, with between 4 and 9 possible choice alternatives on each, for 2213 total lines of data). Their independent variables are *Ideology*, *Equity*, *Expertise*, and *Efficiency*, plus a *CJ Dummy* of the potential assignee (these are all choice-specific variables), plus *Consensus*, *MWC*, *Importance* and *End-of-Term* variables (which are case-specific).

Their analysis points out something else: that you don't necessarily need to interact the observation-specific variables with dummies for the choices...

- So long as they're interacted with *some* variable that varies across alternatives, the model will be identified.

- Here, the theory (such as it is) suggested that they should interact a number of their variables with *ideology*...

Thursday: IIA, MNP, MXL and HEV...

---

[1]M&W call it a "multinomial logit" model and a "discrete choice" model, but it's a conditional logit, with interaction terms to take into account case-specific variables...