MLE: Categorical and Limited Dependent Variables
Unit 2: Ordered, Multinomial, Count, and Limited Dependent Variables
2b. Conditional Logit

PS2730-2021

Week 6

Professor Steven Finkel

- Thus far have considered effect of **individual-specific** factors (distrust, maleness, age, partisanship) on choice of three different outcomes.
- Important aspect of this kind of set-up is that, for three category outcome, we have **one** value of the IV per individual on each variable and we get **two** different coefficients, effect of that variable on log-odds of choice 1 over 3, and effect of that variable on log-odds of choice 2 over 3. So the "effect" of the variable differs depending on the choice comparison.
- But may be other kinds of factors that influence individual utility (or whatever process leads to a given choice).
- For example, choosing a graduate school for PhD studies might depend on distance from home country, how difficult the program is, strength in your area of specialization, etc, as well as whether you have an MA or not, your GRE scores, your career goals, etc. These variables are what are known as **"choice-specific" or "alternative-specific,"** in that they **vary for each choice for each individual**.

- Example: for a person who comes from Berlin, Pitt is 4200 miles away, LSE is 685 miles away, and Hertie School in Berlin is 1 mile away,
- For a second person, who comes from Cleveland, Pitt is 130 miles away, LSE is 3700 miles away, and the Hertie School is 4200 miles.
- **So the attribute of the choice varies by choice by individual.**

- Another example: Effect of the length of commute on choice of bus, subway, or drive to work. Some people's bus rides will be short, some long, some shorter than subway, some slower than driving, some longer, etc. That is **alternative-specific** for a given individual

- Household income, age, partisanship: all **individual-specific** variables
- In our 1992 election example: all we have looked at so far are individual-specific variables. But what about effect of **ideology** on vote, specifically, how far each candidate is from a voter's ideology?
- Person A: really close to Bush ideologically, far from Clinton, moderate to Perot; Person B: really far from Bush, far from Clinton, moderately close to Perot; Person C: moderately far from Bush, close to Clinton, far from Perot

- **Can see this as a perfect example of an alternative-specific variable. As "ideological closeness" increases, vote probability may increase, but ideological closeness varies for each outcome (B,C,P) for each individual.**

- What we have is situation where we have 3 different values for X for each individual and we want to get one β – Notice: As ideological closeness increases, vote probability for a given candidate increases by some function β – but there are three different ideological closeness values for each individual.
- This is the **Conditional Logit Model**, predicting the choice of outcome from a series of alternative-specific variables for each individual
- Important causal processes for choice, but unfortunately under-utilized in political science. Probable reason: **You cannot include choice-specific variables within the traditional MNL set-up.**
- Need to restructure the data and change the model slightly in order to estimate. **BUT: all traditional MNL models can also be estimated within the restructured Conditional Logit format.**
- So Conditional Logit is a more general model, can encompass both individual and alternative-specific variables.

- Conditional Logit set-up: let **Z** be alternative-specific variables (equivalent of **X**), and let γ represent the regression coefficient of **Z** (equivalent of β)
- So Z equals ideological distance from particular candidate (B,C,P), and γ represents the effect of distance on choice

$$\Pr(y = m \mid z) = \frac{\exp^{(Z_m \gamma)}}{\sum_{j=1}^{J} \exp^{(Z_j \gamma)}}$$

$$\Pr(y = m \mid z) = \frac{\exp^{(Z_m \gamma)}}{\sum_{j=1}^{J} \exp^{(Z_j \gamma)}}$$

- So the Prob of Clinton, given Z, is e^distance from Clinton*γ, divided by e^distance from Bush* γ plus e^distance from Clinton* γ + e^distance from Perot*γ
- Similar to MNL in form -- but in that set-up we had constant **X** and different β corresponding to different outcomes. Here we have variable **Z** but the same **γ**, yielding different probabilities depending on the value of Z for each choice for each individual

- We can express the model, as in all logit-type models, in odds terms:

$$\Pr(y = m \mid z) = \frac{\exp^{(Z_m \gamma)}}{\sum_{j=1}^{J} \exp^{(Z_j \gamma)}}$$

$$\Pr(y = n \mid z) = \frac{\exp^{(Z_n \gamma)}}{\sum_{j=1}^{J} \exp^{(Z_j \gamma)}}$$

$$Odds(m \mid n) = \frac{\Pr(y = m \mid z)}{\Pr(y = n \mid z)} = \frac{\dfrac{\exp^{(Z_m \gamma)}}{\sum_{j=1}^{J} \exp^{(Z_j \gamma)}}}{\dfrac{\exp^{(Z_n \gamma)}}{\sum_{j=1}^{J} \exp^{(Z_j \gamma)}}} = \frac{\exp^{(Z_m \gamma)}}{\exp^{(Z_n \gamma)}}$$

- So odds of category m versus n equals the ratio of their exponentiated values on Z multiplied by the (common) regression coefficient

- And in log-odds form

$$LnOdds(m|n) = \ln\left(\frac{\exp^{(Z_m\gamma)}}{\exp^{(Z_n\gamma)}}\right) = Z_m\gamma - Z_n\gamma = (Z_m - Z_n)\gamma$$

- So, for example, if the difference between Clinton ideological distance and Bush ideological distance is 5 units closer to Clinton, then the log-odds of voting for Clinton over Bush will be 5* **γ** greater
- Contrast with Ln-Odds (M|N) in MNL

$$LnOdds(\mathrm{m|n}) = X(B_m - B_n)$$

where one B_j is set to 0 for identification purposes

- One model (MNL) has the difference in regression coefficients for the given category outcome multiplied by a constant X; the other (CL) has the constant regression coefficient multiplied by difference in values of Z for the given categories

Data Set-up for Conditional Logit

- Here is data set up for MNL for first 5 cases:
- NOTE: DISTANCES ARE ((Resp Ideo-Cand Ideo)**2)*-1, SO LESS NEGATIVE NUMBERS MEANS CLOSER TO THE CANDIDATE, MORE NEGATIVE NUMBERS MEANS FARTHER AWAY)

	presvote	bdist	cdist	pdist	distrust
1.	bush	-1	-16	-4	3
2.	bush	-1	-16	-4	4.5
3.	clinton	-4	0	-4	3
4.	bush	-1	-9	-1	4
5.	clinton	-16	-1	-4	5
7.	perot	-4	-1	-9	5

- Cannot use the ideological distance measure in any meaningful way to correspond to CL idea. We have three different values per individual and would get three different coefficients but not in sensical way – what would it mean to say that as distance to Perot gets higher, Prob of Clinton v. Bush gets higher, controlling for distance to Clinton and distance to Bush?

- Strange Model. What we want is a set up that gives one ideological distance measure for each candidate and that becomes the only thing that is considered **for that candidate**. So we want a set-up that has, for each individual, 3 different rows of data, one corresponding to Bush, Clinton, Perot, and the ideological distance to each candidate is included as the IV for that row.

	presvote	Outcome(Alt)	IDEODIST	CHOICE
Case 1	bush	BUSH	-1	1
Case 1	bush	CLINTON	-16	0
Case 1	bush	PEROT	-4	0
Case 2	bush	BUSH	-1	1
Case 2	bush	CLINTON	-16	0
Case 2	bush	PEROT	-4	0
Case 3	clinton	BUSH	-4	0
Case 3	clinton	CLINTON	0	1
Case 3	clinton	PEROT	-4	0
Case 4	bush	BUSH	-1	1
Case 4	bush	CLINTON	-9	0
Case 4	bush	PEROT	-1	0
Case 5	clinton	BUSH	-16	0
Case 5	clinton	CLINTON	-1	1
Case 5	clinton	PEROT	-4	0
Case 7	perot	BUSH	-4	0
Case 7	perot	CLINTON	-1	0
Case 7	perot	PEROT	-9	1

First three rows are person #1, who voted for Bush and whose distance to Bush was -1, to Clinton -16, and to Perot -4. Next three rows are person 2, etc.

- So the CL set-up has the impact of IDEODIST on the CHOICE of that category – Do higher (less negative) numbers on IDEODIST make it more likely that the person chooses the outcome from that row versus the two other outcomes? We will get predicted P for choosing that outcome based on IDEODIST, and the predicted Ps will sum to 1 for each individual over the three outcomes

```
. clogit choice clinton perot ideodist, group(id)

Iteration 0:   log likelihood = -1395.8967
Iteration 1:   log likelihood = -1359.6448
Iteration 2:   log likelihood = -1357.7866
Iteration 3:   log likelihood = -1357.7842
Iteration 4:   log likelihood = -1357.7842

Conditional (fixed-effects) logistic regression

                                                Number of obs     =      4,974
                                                LR chi2(3)        =     927.43
                                                Prob > chi2       =     0.0000
Log likelihood = -1357.7842                     Pseudo R2         =     0.2546

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     clinton |   .4433092   .0665393     6.66   0.000     .3128945    .5737238
       perot |  -.6337188   .0748734    -8.46   0.000    -.7804679   -.4869698
    ideodist |    .142924   .0070546    20.26   0.000     .1290972    .1567509
------------------------------------------------------------------------------
```

- As IDEODIST for candidate 1 relative to candidate 2 increases by 1 unit, the log-odds of choosing that candidate increases by .143, controlling for base log-odds of voting for Clinton and Perot relative to Bush

Combining MNL and CL

- As noted, nice feature of CL is that it can subsume individual-specific variables within the model, so you can have variables that are constant within units as well as alternative-specific variables in a more general choice model
- Data set-up needs to be modified further from what we've seen so far

EXPANDED CL SET-UP

	presvote	Outcome	IDEODIST	CHOICE	CLINTON	PEROT	DISTRUST	TRUSTC	TRUSTP
Case 1	bush	BUSH	-1	1	0	0	3	0	0
Case 1	bush	CLINTON	-16	0	1	0	3	3	0
Case 1	bush	PEROT	-4	0	0	1	3	0	3
Case 2	bush	BUSH	-1	1	0	0	4.5	0	0
Case 2	bush	CLINTON	-16	0	1	0	4.5	4.5	0
Case 2	bush	PEROT	-4	0	0	1	4.5	0	4.5

- Can't just add the individual-specific variable "distrust" to the CL model because there is **no variation within cases** and so it won't help to predict whether "CHOICE" is row 1, row 2 or row 3. So whether the individual unit is high or low on "distrust" has a zero relationship with choice.

- What we can do, though, is construct **interaction terms** of the dummy variable that stands for the category for Clinton and the category for Perot with each individual-specific variable. So we have a variable that is TRUST*CLINTON and TRUST*PEROT, with Bush still being the baseline.

	presvote	Outcome	IDEODIST	CHOICE	CLINTON	PEROT	DISTRUST	TRUSTC	TRUSTP
Case 1	bush	BUSH	-1	1	0	0	3	0	0
Case 1	bush	CLINTON	-16	0	1	0	3	3	0
Case 1	bush	PEROT	-4	0	0	1	3	0	3
Case 5	clinton	BUSH	-16	0	0	0	5	0	0
Case 5	clinton	CLINTON	-1	1	1	0	5	5	0
Case 5	clinton	PEROT	-4	0	0	1	5	0	5
Case 7	perot	BUSH	-4	0	0	0	5	0	0
Case 7	perot	CLINTON	-1	0	1	0	5	5	0
Case 7	perot	PEROT	-9	1	0	1	5	0	5

- So the model is: when distrust is high in the Clinton category, does that lead to a higher probability of CHOICE for the Clinton row being 1, compared to Bush? When distrust is high in the Perot category, does that lead to a higher probability of CHOICE for the Perot row being 1, compared to Bush? So we are always modeling whether that row has a choice of 1 or not, and these interactions give you the effect of distrust on one category over the other, just like in MNL!!

- Can see this in equation form: $$\Pr(y=m\,|\,z)=\frac{\exp^{(Z_m\gamma)}}{\sum_{j=1}^{J}\exp^{(Z_j\gamma)}}$$

where, in our example, z_1 would be Trust*Clinton, and z_2 would be Trust*Perot, z_3 would be a dummy variable for the Clinton row, and z_4 would be a dummy variable for the Perot row

$$\Pr(\text{Clinton}\,|\,\text{z})=\frac{\exp^{(TrustC*\gamma_{dc}+Clinton*\gamma_c)}}{\exp^{(TrustC*\gamma_{dc}+Clinton*\gamma_c)}+\exp^{(TrustP*\gamma_{dp}+Perot*\gamma_p)}+\exp^{(TrustB*\gamma_{db}+Bush*\gamma_b)}}$$

And since the Bush coefficients are all 0 for identification, we get

$$\Pr(\text{Clinton}\,|\,\text{z})=\frac{\exp^{(TrustC*\gamma_{dc}+Clinton*\gamma_c)}}{\exp^{(TrustC*\gamma_{dc}+Clinton*\gamma_c)}+\exp^{(TrustP*\gamma_{dp}+Perot*\gamma_p)}+1}$$

$$\Pr(\text{Perot}\,|\,\text{z})=\frac{\exp^{(TrustP*\gamma_{dc}+Perot*\gamma_c)}}{\exp^{(TrustC*\gamma_{dc}+Clinton*\gamma_c)}+\exp^{(TrustP*\gamma_{dp}+Perot*\gamma_p)}+1}$$

$$\Pr(\text{Bush}|\,\text{z})=\frac{1}{\exp^{(TrustC*\gamma_{dc}+Clinton*\gamma_c)}+\exp^{(TrustP*\gamma_{dp}+Perot*\gamma_p)}+1}$$

$$\Pr(\text{Clinton} \mid z) = \frac{\exp^{(TrustC^*\gamma_{dc} + Clinton^*\gamma_c)}}{\exp^{(TrustC^*\gamma_{dc} + Clinton^*\gamma_c)} + \exp^{(TrustP^*\gamma_{dp} + Perot^*\gamma_p)} + 1}$$

$$\Pr(\text{Perot} \mid z) = \frac{\exp^{(TrustP^*\gamma_{dc} + Perot^*\gamma_c)}}{\exp^{(TrustC^*\gamma_{dc} + Clinton^*\gamma_c)} + \exp^{(TrustP^*\gamma_{dp} + Perot^*\gamma_p)} + 1}$$

$$\Pr(\text{Bush} \mid z) = \frac{1}{\exp^{(TrustC^*\gamma_{dc} + Clinton^*\gamma_c)} + \exp^{(TrustP^*\gamma_{dp} + Perot^*\gamma_p)} + 1}$$

- This is exactly the MNL model!!!
- The coefficient for the dummy variable for Clinton in the CL is the dummy effect of Clinton over Bush in MNL, the regression of TRUSTC in CL being the effect of TRUST on Clinton over BUSH in MNL, the coefficient for the dummy variable for PEROT in CL being the dummy effect of Perot over Bush in MNL, and the regression of TRUSTP in CL being the effect of TRUST on Perot over Bush in MNL

- We can combine MNL and CL into one general model as:

$$\Pr(y = m \mid z) = \frac{\exp^{(Z_m\gamma + X_i\beta_m)}}{\sum_{j=1}^{J} \exp^{(Z_j\gamma + X_i\beta_m)}}$$

where the Zs are alternative-specific variables and the X are individual-specific (interaction) variables with the given alternative (row)

- Odds formulation:

$$Odds(m \mid n) = \frac{\exp^{(Z_m\gamma + X_i\beta_m)}}{\exp^{(Z_n\gamma + X_i\beta_n)}}$$

- Stata implements this model with "clogit" with manually-constructed interaction terms, or, in V17 within the "cm" (choice models) module as "cmclogit" with the interaction terms being constructed automatically via the "casevars" options