UNIVERSITY OF MANNHEIM
School of Social Sciences

Advanced Quantitative Methods in Political Science: Conditional Logit Model

Thomas Gschwend | Oliver Rittmann | Viktoriia Semenova

Week 10 - 4 May 2022

Introduction

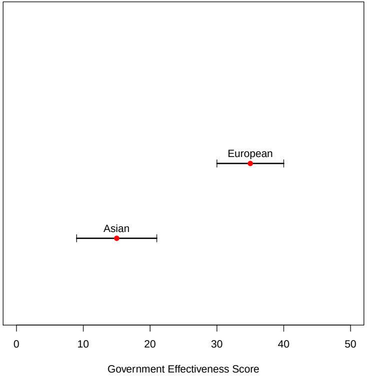
What should you take home from this class today?

- We will broaden the applicability of the *multinominal logit* model (MNL) through a theoretically richer parameterization of the model's systematic component to arrive at the so-called *conditional logit model* (CL)
- We will discuss our simulation strategy to account for *estimation* and *fundamental uncertainty* when calculating quantities of interest.
- Guess what? We will see an example that our discipline can actually do much better than merely $X\beta$ in terms of theory testing.
- We also discuss the so-called *IIA assumption* for MNL and CL models.

Quiz

Quiz

For a sample of countries using MLE we estimate the expected government effectiveness score (i.e., a predicted probability) and the respective 95% CIs for governments around the world. Can we conclude that European governments are on average more effective than Asian governments in our sample?



1. Yes, since the point estimate for Europe is greater than for Asia.
2. Yes, if and only if the confidence intervals do not overlap, the respective difference is systematic.
3. No, because the difference in means is not evaluated.
4. No, because the vertical axis has no measurements.

Conditional Logit Model

Changing the systematic component to derive the Conditional Logit Model

- The systematic component of a decision-maker's utility function might not only depend on her characteristics but also on characteristics of the alternatives (i.e., the different choices).
  - Interesting party characteristics of the *Absurdistan* vote-choice example might be, for instance, size, governmental experience, campaign expenditure, ...
  - Such variables vary across alternatives j (not only across all decision-makers i).
- Take the familiar spatial model, where a decision-maker's utility depends on the distance in a policy space between her and the (perceived) positions of the parties in her choice-set. Note, such issue-distance variables vary across alternatives j *and* decision-makers i as well.

The Conditional Logit Model (CL)

- Thus, as before in the MNL case, let U_{ij} be the utility of decision-maker i derived when choosing alternative $j \in \{1, \ldots, J\}$.
- Given that we cannot observe the decision-maker's utility, we need to specify the *systematic component* such that it relates observable characteristics of the decision-maker (X_i) as well as characteristics of the alternatives (Z_{ij}) to her utility $U_{ij} = V_{ij} + \epsilon_{ij}$. Thus,

$$V_{ij} = X_i\beta_j + Z_{ij}\gamma$$

The Conditional Logit Model (CL)

- Thus, the probability that decision-maker i chooses alternative j is:

$$Pr(y_i = j|X_i, Z_{ij}) = \frac{exp(V_{ij})}{\sum_{j=1}^{J} exp(V_{ij})} = \frac{exp(X_i\beta_j + Z_{ij}\gamma)}{\sum_{j=1}^{J} exp(X_i\beta_j + Z_{ij}\gamma)},$$

  with $\beta_1 = (\beta_{11}, \beta_{12}, \ldots, \beta_{1(k+1)})' = 0$
- Note that in terms of the systematic component V_{ij}, we can write the probability that decision-maker i chooses alternative j in exactly the same terms.
- Multinomial and conditional logit models are variants of the same model. They only have different parameterized systematic components V_{ij}.

Deriving the Likelihood Function for CL

Given that each decision-maker chooses one and only one alternative we use an indicator y_{ij} to accomplish this. Thus, as with the MNL before, the log-likelihood contribution L_i of decision-maker i is

$$\begin{aligned} lnL_i &= ln(\prod_{j=1}^{J} Pr(y_i = j)^{y_{ij}}) \\ &= \sum_{j=1}^{J} y_{ij} ln(Pr(y_i = j)) \end{aligned}$$

Then summing-up all N individual contributions assuming independent realizations gives us the log-likelihood of the conditional logit model.

$$\begin{aligned} lnL(\beta_2, \ldots, \beta_J, \gamma) &= \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} ln(Pr(y_i = j)) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} ln(\frac{exp(X_i\beta_j + Z_{ij}\gamma)}{\sum_{m=1}^{J} exp(X_i\beta_m + Z_{im}\gamma)}) \end{aligned}$$

Chooser- and Choice-specific Data

Again, suppose we are interested in explaining vote-choice in *Absurdistan* with its stable 3-party-system consisting of $1 =$ the Blues, $2 =$ the Reds and $3 =$ the Greens.

- Suppose the *systematic component* of our model consists of chooser characteristics *Age* and *Education* as well as a choice-characteristic *IssueDistance*.
- Thus: $V_{ij} = X_i\beta_j + Z_{ij}\gamma = \beta_{j1} + \beta_{j2} \cdot Age_i + \beta_{j3} \cdot Education_i + \gamma_1 \cdot IssueDistance_{ij}$
- Given that the choice-set consists of three parties, the *systematic components* for decision-maker i with $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})' = 0$ are:

$$
\begin{aligned}
V_{i1} &= \beta_{11} + \beta_{12} \cdot Age_i + \beta_{13} \cdot Education_i + \gamma_1 \cdot IssueDistance_{i1} \\
V_{i2} &= \beta_{21} + \beta_{22} \cdot Age_i + \beta_{23} \cdot Education_i + \gamma_1 \cdot IssueDistance_{i2} \\
V_{i3} &= \beta_{31} + \beta_{32} \cdot Age_i + \beta_{33} \cdot Education_i + \gamma_1 \cdot IssueDistance_{i3}
\end{aligned}
$$

- As you can see, the effect (β's) of our chooser-specific variables vary across the three alternatives while the effect (γ_1) of our choice-specific variable does not.

CL needs data in long-format

			DV	X_1	X_2	Z_1
RespID	Vote	Party	**Choice**	Age	Education	IssueDistance
1	2	1	**0**	18	1	3
1	2	2	**1**	18	1	1
1	2	3	**0**	18	1	1
2	1	1	**1**	25	2	2
2	1	2	**0**	25	2	0
2	1	3	**0**	25	2	1
3	1	1	**1**	27	2	0
3	1	2	**0**	27	2	2
3	1	3	**0**	27	2	2
4	3	1	**0**	43	3	1
4	3	2	**0**	43	3	0
4	3	3	**1**	43	3	3
5	1	1	**1**	54	1	1
5	1	3	**0**	54	1	2
6	3	2	**0**	73	3	3
6	3	3	**1**	73	3	0

A Table of CL estimation results

Table 3: Conditional logit model results predicting behavioral personalization

	Behavioral Personalization			
Campaign Intensity	-0.03*			
	(0.02)			
Campaign Norm	-0.11*			
	(0.06)			
Campaign Content	0.31			
	(0.20)			
Campaign Organization	0.05***			
	(0.02)			
Exposure	-0.02			
	(0.27)			
Partisanship	0.68***			
	(0.08)			
Cognitive Personalization	1.71***			
	(0.39)			
	SPD	Greens	FDP	Left
	Reference Party: CDU/CSU			
East	-1.30	-1.32	-1.17	-1.07
	(0.89)	(0.83)	(0.91)	(0.95)
Political Knowledge	0.02	0.27	0.11	-0.89*
	(0.46)	(0.47)	(0.51)	(0.52)
Intercept	0.52	0.55	0.47	1.61
	(0.73)	(0.86)	(0.81)	(0.99)
Log-Likelihood	-114			
No. of voter-candidate dyads	799			
No. of voters	285			

Note: Entries are estimated coefficients with standard errors in parentheses.
* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. CDU/CSU is reference category.

Interpretation of the CL Estimation Results

Calculate Quantities of Interest

Take a look at the code we used for logit/probit. We just have a different systematic components (including the normalization). Thus, ...

- predicted probabilities.
- first differences.
- ...

You can estimate this model in `R` for instance using the `library(mlogit)`.

Modifying our R-Code

```
# =========================================
# = Step 3: Get theta_c (linear predictor)  =
# =========================================

#INSTEAD OF .... #  theta_1 <- S %*% oil
                 #     ev1  <- 1/(1+exp(-theta_1))
                 # theta_2  <- S %*% nonoil
                 #     ev2  <- 1/(1+exp(-theta_2))

# USE THE FOLLOWING .....
theta_b <- S %*% blue   # make sure scenario is well-defined
theta_r <- S %*% red    # because some coeficients in S have to be set
theta_g <- S %*% green  # to zero! (beta_1k for baseline among others)

ev_blue  <- exp(theta_b)/(exp(theta_b) + exp(theta_r) + exp(theta_g))
ev_red   <- exp(theta_r)/(exp(theta_b) + exp(theta_r) + exp(theta_g))
ev_green <- exp(theta_g)/(exp(theta_b) + exp(theta_r) + exp(theta_g))
```

OK, lets do this step-by-step

```r
# Read-in the data and estimate a conditional logit model
data <- read.dta("absurdistan.dta", convert.factors = FALSE)

ABvote <- mlogit.data(data, shape = "long", choice = "choice",
                      alt.var = "party", id = "respid")

m <- mlogit(choice ~ disLR | educ + age , data = ABvote)
summary(m)

coef(m); vcov(m)
```

Estimation Results of a CL Model

```
> summary(m)

Coefficients :
           Estimate Std. Error  t-value  Pr(>|t|)
alt2      -0.807542   0.287254  -2.8112 0.0049350 **
alt3      -2.864844   0.362069  -7.9124 2.442e-15 ***
disLR     -0.729309   0.037649 -19.3710 < 2.2e-16 ***
alt2:educ  0.361627   0.098887   3.6570 0.0002552 ***
alt3:educ  0.831419   0.120160   6.9193 4.540e-12 ***
alt2:age  -0.485551   0.210337  -2.3084 0.0209744 *
alt3:age  -0.472063   0.249504  -1.8920 0.0584906 .
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

Log-Likelihood: -616.98
McFadden R^2:  0.41591
Likelihood ratio test : chisq = 878.64 (p.value=< 2.22e-16)
```

As Before: Simulate Estimation Uncertainty

```
# ========================================
# = Step 1: Simulate Estimation Uncertainty =
# ========================================


# For MVRNormal we need two parameters (mean and variance)
# get gamma and V from mlogit
gamma <- m$coefficients      # estimated betas, gammas
    V <- solve(-m$hessian)   # inverted neg. Hessian

library(MASS)                                  # provides mvrnorm

nsim <- 10000                                  # N simulations
set.seed(1234)                                 # to replicate draws
S <- mvrnorm(nsim, mu = gamma, Sigma = V) # simulations of gamma
```

Define Scenario of Interest in terms of S

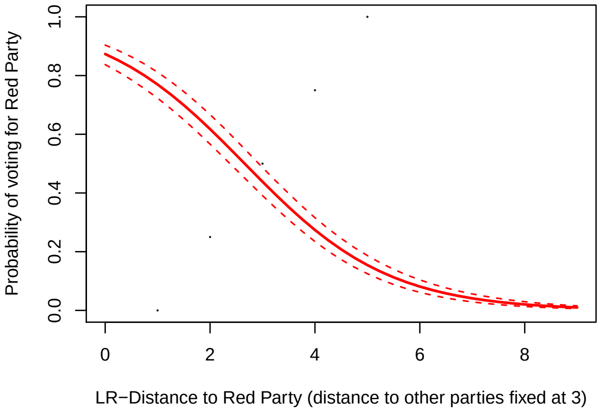
```
# ========================================
# = Step 2: Choose Scenario of Interest  =
# ========================================
> head(S)
        alt2    alt3    disLR  alt2:educ alt3:educ alt2:age alt3:age
[1,] -0.2942 -2.7003 -0.75341 0.331169 0.87911 -1.17044 -0.702439
[2,] -0.6990 -3.0388 -0.72884 0.309437 0.84576 -0.38923 -0.280940

# Given S the corresponding scenario (vector)
# is structured as follows:
# c(alt2, alt3, disLR, alt2:educ, alt3:educ, alt2:age, alt3:age)

red    <- c(1, 0, mean(ABvote[ABvote$party==2,"disLR"]),
           mean(ABvote[,"educ"]), 0, mean(ABvote[,"age"]), 0)
green  <- c(0, 1, mean(ABvote[ABvote$party==3,"disLR"]), 0,
           mean(ABvote[,"educ"]), 0, mean(ABvote[,"age"]))
blue   <- c(0,0,mean(ABvote[ABvote$party==1,"disLR"]), 0, 0, 0, 0)
```

Back to where we started

```
# ==========================================
# = Step 3: Get theta_c (linear predictor)  =
# ==========================================

theta_b <- S %*% blue    # make sure scenario is well-defined
theta_r <- S %*% red     # because some coeficients in S have to be set
theta_g <- S %*% green   # to zero! (beta_1k for baseline among others)

ev_blue  <- exp(theta_b)/(exp(theta_b) + exp(theta_r) + exp(theta_g))
ev_red   <- exp(theta_r)/(exp(theta_b) + exp(theta_r) + exp(theta_g))
ev_green <- exp(theta_g)/(exp(theta_b) + exp(theta_r) + exp(theta_g))

mean(ev_blue); mean(ev_red); mean(ev_green)
quantile(ev_blue, c(.025,.975))
```

When $E(Y_c) = \theta_c$, one can skip step 4 and 5. Once we have π_i, we do not need to draw Y_c and than average them (to incorporate *fundamental uncertainty*) to get $E(Y_c) = \pi_i$.

… and we can even plot them



For more details about the R-Code see `class10.R`

CL - Varying Choice Set

- One can generalize the conditional logit model to allow for varying choice-sets across decision-makers. Examples:
  - Parties might not run in every district (length of ballot varies).
  - Which coalition will form (depends on no. of parties)?
  - In *Judicial Politics*, e.g., the probabilities associated with assigning the opinion in case i to justice j. Choice-set reflects varying majorities across cases.
  - Biostatisticians and epidemiologists call this a $(k_{1i} : k_{0i})$*-matched case–control design*, where $J_i = k_{1i} + k_{0i}$ is the size of the choice set for respondent i (e.g., total no. of alternatives in group i).
  - Economists call this model also a *fixed-effects logit model* for *unbalanced panel data*, or *cross-sectional time-series data* to model war/peace in each country-year or individual turnout in a series of elections (more than one '1' per group possible!).
- Then, the LL is calculated relative to the number of observed choices in the choice-set J_i of the i^{th} decision-maker (i.e., sum over J_i instead of J in the denominator of the LL).
- The model estimates the probability of observing a 1 conditional on the total number of ones (i.e., $\#1$) and the number of choices J_i for each decision-maker (a conditional likelihood is computed recursively).

How we can do better than merely using $X\beta$ for theory testing

Empirical Implications of Theoretical Models (EITM)

- You can use a conditional logit model (or any other statistical model) to estimate a utility function of a decision-maker that your theory actually comes up with.
- Advantage: You can parametrize (and finally estimate) the systematic component V_{ij} of your utility function according to your theory.

Example: Comprehensive Theory of Party-Choice in Absurdistan

- Example: Suppose you develop a more comprehensive theory of party-choice in Absurdistan.
- The basic components of your theory are that party-choice depends on policy as well as non-policy factors. Hence,

$$V_{ij} = \alpha(Policy) + \gamma(Non - Policy)$$

- This is well known from the literature. Your contribution, though, is to test whether voters are more likely to behave according to the directional or the proximity theory of how policy factors have an impact on vote-choice. Hence,

$$V_{ij} = \alpha\{(1-\beta)\cdot(Directional) + \beta\cdot(Proximity)\} + \gamma(Non - Policy)$$

- Let x_i be voter i's position on the LR-dimension and p_j the position of party j on the same dimension measured on a 7-pt scale (conceptually better would be to take a voter's perceived position p_{ij} of party j). Hence,

$$V_{ij} = \alpha\{(1-\beta)\cdot[(x_i-4)(p_j-4)] + \beta\cdot(x_i-p_j)^2\} + \gamma(Non - Policy)$$

Implementation in R - for details see PartyChoiceAbsurdistan.R

```
Lik_mixedmodel <- function(theta
                           ,Z # response matrix N*K
                                # True if voter selects party
                                # else FALSE
                           ,c # vector party positions
                           ,x # vector voter position
                           ,pid # matrix partyidentification
                           ){
J <- ncol(Z) # Number of parties
N <- nrow(Z) # Number of voters
V <- matrix(NA, ncol=J,nrow=N)  # Utilities (systematic component)

for (j in 1:J){ # Different parties
V[,j] <-  theta[1]*((1-theta[2])*(x-4)*(c[j]-4)+theta[2]*((x-c[j])^2))+theta[3]*pid[,j]
 }

# Note,  theta[1] = \alpha, theta[2] = \beta, theta[3] = \gamma

# Create matrix for choice probabilities & vector for linear predictor
P <- matrix(NA, ncol=J, nrow=N)
expsum  <- NULL

# Calculate sum for each individual;
for (i in 1:N) expsum[i] <- sum(exp(V[i,]))
for (j in 1:J) P[,j] <-  exp(V[,j])/expsum

# Return Log liklihood
return(sum(log(P[Z])))
}
```

Limitations of MNL or CL Models

Limitations of MNL or CL Models

- Such logit models can account for *systematic taste variation* (the value decision-makers place on each characteristic of the alternatives) but not for random taste variation due to unobserved factors (i.e., unobserved heterogeneity).
  - One can include interaction effects between a choice-specific variable and a chooser-specific variable to capture the idea that an alternative characteristic is more important for some than for others.
  - Nevertheless, two voters with the same values on all independent variables might make different choices because of unobserved factors.
  - Include a random effect to account for unobserved heterogeneity (*random parameter logit* can be estimated in R)
- If unobserved factors are *not* independent over time in repeated choice situations such that they introduce correlations over time, use so-called *discrete time duration models*.

A Major Limitation? The IIA Assumption

- The ratio of logit probabilities for any two alternatives j and k is:

$$\frac{Pr(y=j)}{Pr(y=k)} = \frac{exp(V_{ij})}{exp(V_{ik})} = e^{V_{ij}-V_{ik}}$$

  - The key thing to note is that the above ratio does only depend on alternatives j and k.
- Because the ratio is independent of alternatives other than j and k, those logit models are said to be *independent of irrelevant alternatives* (IIA)
- IIA implies that the presence or absence of another alternative (e.g. strategic entry of candidates or parties) should not alter the relative probabilities of any single decision-maker (conditional on the model's systematic component!). If alternatives are seen as substitutes then adjust the systematic component to account for that.
- Logit models assume IIA and cannot capture more flexible substitution patterns across alternatives. Use *Multinomial Probit* or *Nested Logit* models if needed, although they also have serious drawbacks (e.g., computationally hard, weakly identified, defending nesting-structure)

Testing for IIA Assumption

- There are two formal tests (*Hausman-McFadden Test, Small-Hsiao Test*). However, they often yield inconsistent and conflicting results.
    - IIA was never a big deal in my applications.
    - Especially, if the choice situation is modeled realistically.
- Eyeballing (or a *Eyeball Test* - if you will).
    - One way to think about IIA is that the estimated coefficients should be invariant to which choices are available.
    - Drop one (or more) party at a time and reestimate the model. Check whether the coefficients change or not. If nothing changes IIA seems to hold.

Testing for IIA Assumption in R

- We can test IIA using the *Hausman-McFadden test*. We first run two ML or CL models and apply **mlogit** first to all choice outcomes and then (successively) to a subset of choice outcomes. We then use **hmftest** to calculate the test statistic.
- It should *not* matter whether we remove one choice (H_0).
- Hausman, J.A. and D. McFadden. 1984. "A Specification Test for the Multinomial Logit Model", *Econometrica* 52: 1219-40.

```
# for subs party==3 is removed from estimation
full <- mlogit(choice ~ disLR | educ + age , data = ABvote)
subs <- mlogit(choice ~ disLR | educ + age , data = ABvote, alt.subset=c("1", "2"))

hmftest(full, subs)

      Hausman-McFadden test

      data:  ABvote
      chisq = 16.9384, df = 4, p-value = 0.001987
      alternative hypothesis: IIA is rejected
```