PPOL 503-03.PPOL 503-04, Fall 2016
Course Notes #8 Multinomial Logit Model

Discrete dependent variable that takes on more than two outcomes and the outcomes or categories are unordered.

- Examples:
  - Travel mode of urban commuters (bus, train, car)
  - Occupational choice (craft, menial, blue collar, white collar, professional).
- A set of equations, one for each outcome category, is estimated. To identify the parameters of the model, the normalization of one set of estimates=0 (the “base category”) is imposed.

$$P(Y_i = j) = \frac{\exp(\beta_j' X_i)}{\sum_{k=1}^{r} \exp(\beta_k' X_i)}$$

- The estimated equations provide a set of probabilities for the r choices for the decision maker with attributes X.
- The category chosen as "base outcome" is irrelevant. The results yield the same estimated probabilities for each outcome. But the interpretation of the coefficients are in relation to the base category.

- Example: 3 outcome categories.

$$P(Y=1)=\frac{\exp(X\beta^{(1)})}{\exp(X\beta^{(1)})+\exp(X\beta^{(2)})+\exp(X\beta^{(3)})}$$

$$P(Y=2)=\frac{\exp(X\beta^{(2)})}{\exp(X\beta^{(1)})+\exp(X\beta^{(2)})+\exp(X\beta^{(3)})}$$

$$P(Y=3)=\frac{\exp(X\beta^{(3)})}{\exp(X\beta^{(1)})+\exp(X\beta^{(2)})+\exp(X\beta^{(3)})}$$

- Setting category 1 as "base" category:

$$P(Y=1)=\frac{1}{1+\exp(X\beta^{(2)})+\exp(X\beta^{(3)})}$$

$$P(Y=2)=\frac{\exp(X\beta^{(2)}}{1+\exp(X\beta^{(2)})+\exp(X\beta^{(3)})}$$

$$P(Y=3)=\frac{\exp(X\beta^{(3)}}{1+\exp(X\beta^{(2)})+\exp(X\beta^{(3)})}$$

- Relative probability of category 2 to the base (1):

$$\frac{P(Y=2)}{P(Y=1)} = \exp(X\beta^{(2)})$$

- The odds ratio between two alternatives (relative risk of category 2 to category 1) does not depend on the other alternatives: Independence of Irrelevant Alternatives (IIA). This means that adding or deleting alternative outcome categories does not affect the odds among the remaining outcomes. Multinomial probit can be used as an alternative.
- What is the effect of a one-unit change in X?

  The ratio of the relative risk of the two categories evaluated at X+1 and at X:

$$\frac{\exp\left[(X+1)(\beta^{(2)})\right]}{\exp\left[(X)(\beta^{(2)})\right]} = \exp(\beta^{(2)})$$

- Issues in Interpretation of Multinomial Logit.
  - Difficulty in interpretation: numerous coefficients, various equations. Results refer to relative changes (compare alternatives in pairs).
  - The exponentiated value of a coefficient is the relative risk ratio for a one-unit change in the corresponding variable.
  - Each set of coefficients must be interpreted relative to the reference or base category.
  - If the outcome of interest has four choices, the multinomial logit estimation will yield three sets of parameters.

- Independence of Irrelevant Alternatives (IIA)

  This property means that the probability ratio of any two alternatives *j* and *l* does not depend on any alternatives other than *j* and *l*.

  In other words, the ratio of probabilities of any two alternatives is necessarily the same regardless of what other alternatives are in the choice set or what the characteristics of the other alternatives are. Again, adding or deleting alternative outcome categories does not affect the odds among the remaining outcomes.

- IIA assumption: convenient for calculation and forecasting.
- Criticism: restrictive because it imposes constraints on behavior. Need to be careful in how one defines the choice set.
- Example: choice between red and blue bus if you add a yellow bus, it will not affect the probability ratio between choosing a red or blue bus.
- Example: choice between train, car and bus is less clear cut; IIA assumption may be questionable.

- The IIA restriction does not apply to the population as a whole. That is, it does not restrict the shares of the population choosing any two alternatives to be unaffected by the utilities of other alternatives.
- The key to understanding this distinction is that for homogenous market segments IIA holds, but across market segments unobserved attributes vary, and thus the IIA property does not hold for a population of individuals.
- The MNL is an appropriate model if the systematic component of utility accounts for heterogeneity across individuals. In general, models with many socio-economic variables have a better chance of not violating IIA.

- Testing the IIA assumption: use a Hausman and McFadden specification test. Compare the restricted model to the unrestricted model with the null hypothesis that the coefficients of the restricted choice are equal to zero. Reject the null if the Chi-square is highly significant. If a choice (or subset of choices) is irrelevant, then excluding it should not change the parameter estimates. But if the remaining odds-ratios are not truly independent of these alternatives, the parameter estimates obtained when this choice is eliminated will be inconsistent.

$$Chi-Square=(\hat{\beta}_s-\hat{\beta}_f)'\left[\hat{V}_s-\hat{V}_f\right]^{-1}(\hat{\beta}_s-\hat{\beta}_f)$$

where s = restricted subset

f = full set of choices

$\hat{V}_s$, $\hat{V}_f$ = covariance matrices

- Example: Wish to examine factors associated voting behavior during the 1996 presidential election. Alternatives:
  - Vote for Dole (589 respondents)
  - Vote for Clinton (923 respondents)
  - No vote (830 respondents)
- Control for: party identification, political views, race, education and region.
- Results are reported in Table 1. The multinomial logit coefficients in the upper panel show the effects of voting for Dole versus nonvoting, and the lower indicate voting for Clinton relative to nonvoting.

- Multinomial Logit Estimates for the 1996 Presidential Election (Non-voters are reference)

Independent	logit coeff	std. error	Wald Chi	odds ratio
Dole Voters				
Intercept	-8.259	(0.516)	-16.00**	-----
Party ID	0.513	(0.044)	11.66**	1.67
Political View	0.434	(0.056)	7.75**	1.54
White	1.397	(0.251)	5.6**	3.94
Education	0.220	(0.025)	8.80**	1.25
South	-0.100	(0.139)	-.72	N/S
Clinton Voters				
Intercept	-0.434	(0.324)	-1.33	-----
Party ID	-0.484	(0.036)	-13.44**	0.65
Political View	- 0.040	(0.043)	-0.93	N/S
White	0.188	(0.126)	1.49	N/S
Education	0.132	(0.019)	6. 95**	1.14
South	-0.264	(0.114)	-2.32	0.77

- Results in relation to the “base” category (nonvoting).
- Many of the coefficients have opposite signs.
- Party ID = 1 if republican; =0 if democrat
- For example, the +.523 party identification parameter for Dole means that Republican identifiers were more likely to vote for him than to stay home, while the -.484 parameter for Clinton voting means that Republicans were less likely to vote than to cast their ballot for the Democratic candidate.

- Political views = 1 if conservative; =0 if not conservative.
- Similarly, political conservatives were more likely to vote for Dole (.434), but less likely to support Clinton (-.040), than not to vote.
- The positive signs for both sets of voters on race and education indicate that both whites and more educated voters are more likely to vote than to stay home, while the coefficient sizes are somewhat higher for Dole than for Clinton.
- The negative coefficient on region (South = 1) means that Southerners were less likely to vote for either candidate than to stay home.

In the multinomial logit model with ten predictors (two equations with five independent variables)-unrestricted model

$$2 \text{ Log } L_1 = -2{,}228.5$$

For the equation that includes only the intercept (no predictors)—restricted model

$$2 \text{ Log } L_0 = -3{,}489.4$$

The model improvement in fit is - 2,228.5 – (-3,489.4) = 1,260.9 with 10 DOF. This is highly significant at the .001 level of significance for the chi-square with DOF= 10 is 29.6.

ODDS RATIO INTERPRETATION

DOLE VOTERS:

1) PARTY ID: OR = 1.67: Individuals who identified themselves as “republicans” were 1.67 times as likely to vote for Dole than to stay home.

2) VIEWS: OR = 1.54: Individuals who considered themselves to be “conservative” were 1.54 times as likely to vote for Dole than to stay home.

3) RACE- OR = 3.94: Whites were almost 4 times as likely as non-whites to vote for Dole than to stay home.

4. EDUCATION- OR = 1.25; Persons with more than 12 years of schooling were 1.25 times as likely to vote for Dole than to stay home.

5. REGION: not significant

CLINTON VOTERS

1) PARTY ID – OR = .65; Republicans were 65% as likely to vote for Clinton than to stay home.

2) EDUCATION- OR 1.14: Person with more than 12 years of schooling were 1.14 times as likely to vote for Clinton than to stay home.

3) REGION- OR = .77; Residents of the South were 77% as likely to vote for Clinton than to stay home.

4) RACE and VIEWS are not statistically significant