

Homework 6: Discrete Choice Models I

Due Date:

Exercise 1: Comparing Logit and Multinomial Logit

Use nes92.dta, which is a small subset of the 1992 National Election Study (cleaned up) that we saw before. It includes Bush, Clinton, and Perot voters.

1. Start by renaming some variables. `rename v3 CONSERVATIVE; rename v14 EDUCATION; rename v16 UNION; rename v17 INCOME; rename v18 BLACK; rename v1 VOTE; rename v8 DISTANCETOCLINTON; rename v7 DISTANCETOBUSH; rename v10 ECONOMYWORSE;`
2. Estimate a logit model where voters are choosing just between Bush (0) and Clinton (1). The explanatory variables should be: `CONSERVATIVE`, `ECONOMYWORSE`, `EDUCATION`, `UNION`, `INCOME`, `BLACK`.
3. Estimate a logit model where voters are choosing just between Bush (0) and Perot (1). The explanatory variables should be the same.
4. Estimate a logit model where voters are choosing just between Perot (0) and Clinton (1). The explanatory variables should be the same.
5. Now estimate a multinomial logit model where voters are choosing between Bush, Perot, and Clinton. Make Bush the base category.
6. Fill in the information in the table shown on the next page.
7. How and why do the results differ? How and why are the results the same? What is the relationship between the three logit models?

Table 1: The Determinants of Presidential Vote Choice in 1992

Regressor	Reference Candidate: Bush				Reference Candidate: Perot
	Logit Clinton	Logit Perot	Multinomial Clinton	Logit Perot	Logit Clinton
Conservative					
EconomyWorse					
Education					
Union					
Income					
Black					
Constant					
Log likelihood					
Observations					

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed)
(Standard errors are given in parentheses)

Exercise 2: MNL

The next exercise uses Dutch election data. The codebook is `dutch271codebook.txt` (4 main parties only). The STATA data are `dutch271uncond.dta`.

1. Rename the variables: `rename sex MALE`; `rename educ EDUCATION`; `rename abor ABORTION`; `rename nuke NUCLEAR`; `rename inc INCOMEDIFFERENCE`; `rename left7 RIGHT`; `rename relig RELIGIOUS`;
2. Run an MNL model with the `Pvda` as the base category. The independent variables should be `ABORTION` `NUCLEAR` `INCOMEDIFFERENCE` `RIGHT` `MALE` `RELIGIOUS` `EDUCATION`. The dependent variable is `PARTY`.
3. Interpret the coefficients on `ABORTION`, `EDUCATION`, and the `CONSTANT`.
4. Calculate the predicted probability that an individual who is a 2 on `ABORTION`, `NUCLEAR`, `INCOMEDIFFERENCE`, and `RIGHT`, who is `MALE`, who is `RELIGIOUS`, and who is a 5 on `EDUCATION` will vote for the `VVD` rather than the `Pvda`. Be sure to provide 95% confidence intervals (manually through simulation). Interpret.
5. Calculate the predicted probability that an individual with the same characteristics except an `EDUCATION` score of 1 instead of 5 will vote for the `VVD` rather than the `Pvda`. Be sure to provide 95% confidence intervals (manually through simulation). Interpret.
6. Now calculate the change in probability of voting for the `VVD` rather than the `Pvda` associated with changing that individual's `EDUCATION` score from 5 to 1. Again, manually provide confidence intervals. Interpret.
7. Use the code in the notes online to graphically shown a predicted probability line plot for all levels of education for the scenario where the individual is a 2 on `ABORTION`, `NUCLEAR`, `INCOMEDIFFERENCE`, and `RIGHT`, and is `MALE` and `RELIGIOUS`.
8. Use the code in the notes online to graphically shown a summed predicted probability area plot for all levels of education for the scenario where the individual is a 2 on `ABORTION`, `NUCLEAR`, `INCOMEDIFFERENCE`, and `RIGHT`, and is `MALE` and `RELIGIOUS`.
9. Calculate the odds that an individual who is a 2 on `ABORTION`, `NUCLEAR`, `INCOMEDIFFERENCE`, and `RIGHT`, who is `MALE`, who is `RELIGIOUS`, and who is a 5 on `EDUCATION` will vote for the `CDA` rather than the `Pvda`. Be sure to provide 95% confidence intervals (manually through simulation). Interpret.
10. Take the same individual from the previous question. What is the odds of that individual voting for the `VVD` rather than the `Pvda` if we increase their abortion score by one unit? Be sure to provide 95% confidence intervals (manually through simulation). Interpret.
11. What's the odds of voting for the `VVD` rather than the `Pvda` if we increase the abortion score by one unit? Be sure to provide 95% confidence intervals (manually through simulation). Interpret.

Exercise 3: CL

The next exercise uses Dutch election data. The codebook is `dutch271codebook.txt` (4 main parties only). The STATA data are `dutch271cond.dta`.

1. Rename variables: `rename sex MALE`; `rename educ EDUCATION`; `rename disabor ABORTION`; `rename disnuke NUCLEAR`; `rename disinc INCOMEDIFFERENCE`; `rename disleft RIGHT`; `rename relig RELIGIOUS`; `rename abor RESPABOR`; `rename nuke RESPNUKE`; `rename inc RESPINC`; `rename left RESPRIGHT`;
2. Run a pure conditional logit. Use the following independent variables: `ABORTION` `NUCLEAR` `INCOMEDIFFERENCE` `RIGHT`. `CHOICE` is the dependent variable. Interpret the coefficients on `ABORTION` and `NUCLEAR`. How does this analysis differ from using the individual-specific variables in the MNL model?
3. Now run the same model but include the following independent variables: `ABORTION` `NUCLEAR` `INCOMEDIFFERENCE` `RIGHT` `MALE` `RELIGIOUS` `EDUCATION`. What happens and why?
4. Imagine a voter is choosing between two parties, m and n . On `RIGHT`, `INCOMEDIFFERENCE`, and `NUCLEAR`, the two choices are the same distance from the respondent. However, party m is 2 units closer on the `ABORTION` issue than party n . What are the odds that the voter chooses party m over party n ? Provide confidence intervals. Interpret.
5. Drop D66 and do an eyeball test of the IIA assumption. Now drop D66 and PvdA and do an eyeball test of the IIA assumption. What do you find?
6. Now do a series of formal Hausman tests. Fill in the table below:

Table 2: Hausman Test for IIA: Pure Conditional Model

Dropped Parties	χ^2	p-value
PvdA		
CDA		
VVD		
D66		
PvdA & CDA		
PvdA & VVD		
PvdA & D66		
VVD & CDA		
VVD & D66		
CDA & D66		

Exercise 4: Mixed CL

Still using the dutch271cond.dta.

1. Create some individual-specific variables using the dummy variable trick explained in the notes. You should create variables for MALE, RELIGIOUS, EDUCATION, and the CONSTANT.
2. Run a CL model that includes the same individual-specific variables as you did in the original MNL analysis (**except the left-right variable**) i.e. ABORTION, NUCLEAR, MALE, RELIGIOUS, EDUCATION, CONSTANT. Now compare this CL model with the original MNL analysis. What is the same and what is different?
3. Now run a mixed CL model and fill in the following table:

Table 3: The Determinants of Party Choice in the Netherlands

Regressor	Mixed Conditional Logit		
AbortionDistance			
NuclearDistance			
IncomeDifferenceDistance			
RightDistance			
	Reference Party: PvdA		
	CDA	VVD	D66
Male			
Religious			
Education			
Constant			
Log likelihood			
Observations			

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed)
Standard errors are given in parentheses

4. Interpret the coefficients on ABORTIONDISTANCE and MALE.

5. Calculate the odds that an individual will vote for a party m that is closer by one unit on the abortion issue than a party n . Provide confidence intervals and interpret.
6. Calculate the odds that an individual will vote for the VVD rather than the Pvda if he was a male compared to an identical voter who is female. Provide confidence intervals and interpret.
7. Calculate the odds that an individual votes for the CDA rather than the Pvda if the individual is male, religious, has a high level of education (5), when the Pvda is two units further away on the nuclear issue, and when the Pvda and CDA are equally distant on all the other issue dimensions. Provide confidence intervals and interpret.
8. Calculate the predicted probability that an individual votes for the CDA rather than the Pvda if the individual is male, religious, has a high level of education (5). Assume that all of the parties are at the individual's ideal point on the abortion, income, and left-right issue dimensions but that they are two units away on the nuclear issue dimension. Provide confidence intervals and interpret.