Multinomial Models

Week 8
POLS 8830: Advanced Quantitative Methods

Ryan Carlin
Georgia State University
rcarlin@gsu.edu

Theory Behind Multinomial Models

- Can be conceptualized as simultaneously estimating binary logits (probits) for all possible combinations across categories

Theory Behind Multinomial Models

- Can be conceptualized as simultaneously estimating binary logits (probits) for all possible combinations across categories
- With 3 categories, multinomial logit is similar to estimating 3 separate logit equations

Theory Behind Multinomial Models

- Can be conceptualized as simultaneously estimating binary logits (probits) for all possible combinations across categories
- With 3 categories, multinomial logit is similar to estimating 3 separate logit equations
  - Compare outcomes 1-2, 2-3, and 3-1
- Occasionally used to estimate models with ordinal dependent variables

Theory Behind Multinomial Models

- Can be conceptualized as simultaneously estimating binary logits (probits) for all possible combinations across categories
- With 3 categories, multinomial logit is similar to estimating 3 separate logit equations
  - Compare outcomes 1-2, 2-3, and 3-1
- Occasionally used to estimate models with ordinal dependent variables
  - Useful in determining whether dependent variable is truly ordinal

Theory Behind Multinomial Models

- Can be conceptualized as simultaneously estimating binary logits (probits) for all possible combinations across categories
- With 3 categories, multinomial logit is similar to estimating 3 separate logit equations
  - Compare outcomes 1-2, 2-3, and 3-1
- Occasionally used to estimate models with ordinal dependent variables
  - Useful in determining whether dependent variable is truly ordinal
  - Tradeoff involves a loss of efficiency compared to ordered logit because not all information is used in multinomial model (lose the ordering)

Example: Venezuelan Parties

- Suppose a nominal dependent variable tracks three political party choices available to Venezuelan voters:
  - A — Acción Democrática
  - B — Other (i.e. neither Acción Democrática nor COPEI)
  - C — COPEI

Example: Venezuelan Parties

- Suppose a nominal dependent variable tracks three political party choices available to Venezuelan voters:
  - A — Acción Democrática
  - B — Other (i.e. neither Acción Democrática nor COPEI)
  - C — COPEI
- Dataset contains observations across all categories N_A, N_B, and N_c

Example: Venezuelan Parties

- Suppose a nominal dependent variable tracks three political party choices available to Venezuelan voters:
  - A — Acción Democrática
  - B — Other (i.e. neither Acción Democrática nor COPEI)
  - C — COPEI
- Dataset contains observations across all categories N_A, N_B, and N_c
- Also contains a set of independent variables $\mathbf{X}$

Intuition Underlying the Multinomial Logit

- To examine the effects of **X** on the probability of outcome A versus outcome B:

Intuition Underlying the Multinomial Logit

- To examine the effects of **X** on the probability of outcome A versus outcome B:
  - We need to select observations N_A and N_B

Intuition Underlying the Multinomial Logit

- To examine the effects of **X** on the probability of outcome A versus outcome B:
  - We need to select observations N_A and N_B
  - Then estimate a binary logit with *only* those observations

$$\ln\left[\frac{\Pr(A|\mathbf{X})}{\Pr(B|\mathbf{X})}\right] = \beta_{0,A|B} + \beta_{1,A|B}\mathbf{X}$$

Intuition Underlying the Multinomial Logit

- Then we estimate the next pairwise comparison (B and C)

Intuition Underlying the Multinomial Logit

- Then we estimate the next pairwise comparison (B and C)
  - Using *only* observations N_B and N_C

$$\ln\left[\frac{\Pr(B|\mathbf{X})}{\Pr(C|\mathbf{X})}\right] = \beta_{0,B|C} + \beta_{1,B|C}\mathbf{X}$$

- Finally, we estimate the last pairwise comparison (A and C)

Intuition Underlying the Multinomial Logit

- Then we estimate the next pairwise comparison (B and C)
  - Using *only* observations N_B and N_C

$$\ln\left[\frac{\Pr(B|\mathbf{X})}{\Pr(C|\mathbf{X})}\right] = \beta_{0,B|C} + \beta_{1,B|C}\mathbf{X}$$

- Finally, we estimate the last pairwise comparison (A and C)
  - Using *only* observations N_A and N_C

$$\ln\left[\frac{\Pr(A|\mathbf{X})}{\Pr(C|\mathbf{X})}\right] = \beta_{0,A|C} + \beta_{1,A|C}\mathbf{X}$$

Intuition Underlying the Multinomial Logit

- However, do we need to estimate all 3 logit equations?

Intuition Underlying the Multinomial Logit

- However, do we need to estimate all 3 logit equations?
  - If we know how **X** affects the probability of A versus B, and how **X** affects the probability of B versus C, do we not also know how **X** affects the probability of A versus C already?

$$\ln\left[\frac{\Pr(A|\mathbf{X})}{\Pr(B|\mathbf{X})}\right]+\ln\left[\frac{\Pr(B|\mathbf{X})}{\Pr(C|\mathbf{X})}\right]=\ln\left[\frac{\Pr(A|\mathbf{X})}{\Pr(C|\mathbf{X})}\right]$$

Intuition Underlying the Multinomial Logit

- Since the left-hand side of the equations form a linear combination, we can rewrite the right-hand side as well

$$(\beta_{0,A|B} + \beta_{1,A|B}\mathbf{X}) + (\beta_{0,B|C} + \beta_{1,B|C}\mathbf{X}) = (\beta_{0,A|C} + \beta_{1,A|C}\mathbf{X})$$

Intuition Underlying the Multinomial Logit

- This allows us to separately examine the intercept terms and the slope coefficient terms

$$(\beta_{0,A|B}) + (\beta_{0,B|C}) = (\beta_{0,A|C})$$
$$(\beta_{1,A|B}) + (\beta_{1,B|C}) = (\beta_{1,A|C})$$

Intuition Underlying the Multinomial Logit

- In sum the results of the binary logit for A versus C can be derived from the results of the binary logits for A versus B and B versus C

Intuition Underlying the Multinomial Logit

- In sum the results of the binary logit for A versus C can be derived from the results of the binary logits for A versus B and B versus C
- What is the problem here?

Intuition Underlying the Multinomial Logit

- This result is valid only for the population parameters and does not remain valid for the sample estimates

Intuition Underlying the Multinomial Logit

- This result is valid only for the population parameters and does not remain valid for the sample estimates
- The reason involves the use of different observations for the sample estimates

Intuition Underlying the Multinomial Logit

- This result is valid only for the population parameters and does not remain valid for the sample estimates
- The reason involves the use of different observations for the sample estimates
  - Sample one has $N_A + N_B$ observations
  - Sample two has $N_B + N_C$ observations

Intuition Underlying the Multinomial Logit

- This result is valid only for the population parameters and does not remain valid for the sample estimates
- The reason involves the use of different observations for the sample estimates
  - Sample one has $N_A + N_B$ observations
  - Sample two has $N_B + N_C$ observations
  - Therefore, deriving results for a sample with $N_A + N_C$ observations is not possible
- The solution: the multinomial logit model, which estimates the equations simultaneously

Intuition Underlying the Multinomial Logit

- This result is valid only for the population parameters and does not remain valid for the sample estimates
- The reason involves the use of different observations for the sample estimates
  - Sample one has $N_A + N_B$ observations
  - Sample two has $N_B + N_C$ observations
  - Therefore, deriving results for a sample with $N_A + N_C$ observations is not possible
- The solution: the multinomial logit model, which estimates the equations simultaneously
- This approach uses the data more efficiently and does not leave us susceptible to this problem

Mechanics of the Multinomial Logit

- Relies on the logistic distribution

Mechanics of the Multinomial Logit

- Relies on the logistic distribution
- Simultaneously examines the following equations:

$$\frac{\Pr A}{\Pr C} = e^{\mathbf{X}\beta_A}$$

$$\frac{\Pr B}{\Pr C} = e^{\mathbf{X}\beta_B}$$

- Note: One outcome is maintained as a baseline category (in this example C).

Mechanics of the Multinomial Logit

- Since the 3 alternatives together combine to explain all possible outcomes, we can infer the following:

$$\Pr A = \frac{e^{\mathbf{X}\beta_A}}{1 + e^{\mathbf{X}\beta_A} + e^{\mathbf{X}\beta_B}}$$

$$\Pr B = \frac{e^{\mathbf{X}\beta_B}}{1 + e^{\mathbf{X}\beta_A} + e^{\mathbf{X}\beta_B}}$$

$$\Pr C = \frac{1}{1 + e^{\mathbf{X}\beta_A} + e^{\mathbf{X}\beta_B}}$$

Mechanics of the Multinomial Logit

- Therefore the likelihood function becomes:

$$L(\beta_2, \ldots, \beta_J | \mathbf{y}, \mathbf{X}) =$$

$$\prod_i \frac{e^{\mathbf{x}_i \beta_A}}{1 + e^{\mathbf{X}\beta_A} + e^{\mathbf{X}\beta_B}} \prod_j \frac{e^{\mathbf{x}_j \beta_B}}{1 + e^{\mathbf{X}\beta_A} + e^{\mathbf{X}\beta_B}} \prod_k \frac{1}{1 + e^{\mathbf{X}\beta_A} + e^{\mathbf{X}\beta_B}}$$

Mechanics of the Multinomial Logit

- And the log-likelihood becomes:

$$\ln L(\beta_2, \ldots, \beta_J | \mathbf{y}, \mathbf{X}) =$$

$$\sum_i \frac{e^{\mathbf{x}_i \beta_A}}{1 + e^{\mathbf{X}\beta_A} + e^{\mathbf{X}\beta_B}} + \sum_j \frac{e^{\mathbf{x}_j \beta_B}}{1 + e^{\mathbf{X}\beta_A} + e^{\mathbf{X}\beta_B}} + \sum_k \frac{1}{1 + e^{\mathbf{X}\beta_A} + e^{\mathbf{X}\beta_B}}$$

Interpretation of the Multinomial Logit

- Interpretation of coefficients always conducted with respect to the baseline (or comparison) category

Interpretation of the Multinomial Logit

- Interpretation of coefficients always conducted with respect to the baseline (or comparison) category
- This is also true of interpreting marginal effects or predicted probabilities

Interpretation of the Multinomial Logit

- Interpretation of coefficients always conducted with respect to the baseline (or comparison) category
- This is also true of interpreting marginal effects or predicted probabilities
- In the previous example, if "Acción Democrática" is the baseline category, the likelihood of voting for "COPEI" would be interpreted with respect to the baseline likelihood of voting for "Acción Democrática"

Interpretation of the Multinomial Logit

- Interpretation of coefficients always conducted with respect to the baseline (or comparison) category
- This is also true of interpreting marginal effects or predicted probabilities
- In the previous example, if “Acción Democrática” is the baseline category, the likelihood of voting for “COPEI” would be interpreted with respect to the baseline likelihood of voting for “Acción Democrática”
- Similarly, the likelihood of voting “Other” would be interpreted with respect tot he baseline likelihood of “Acción Democrática”

Interpretation of the Multinomial Logit

- By default, `nnet` and `mlogit` use the lowest category (i.e. 1 in an variables of 1, 2, 3) as the baseline

Interpretation of the Multinomial Logit

- By default, `nnet` and `mlogit` use the lowest category (i.e. 1 in an variables of 1, 2, 3) as the baseline
- This is rather atheoretical, so we should always select the most theoretically appropriate category for purposes of comparison (if possible)
    - We use the `relevel()` function

Interpretation of the Multinomial Logit

- By default, `nnet` and `mlogit` use the lowest category (i.e. 1 in an variables of 1, 2, 3) as the baseline
- This is rather atheoretical, so we should always select the most theoretically appropriate category for purposes of comparison (if possible)
    - We use the `relevel()` function
- Moreover, to calculate predictions for the baseline category (regardless of which one is chosen) it must be estimated separately (using a different baseline)

Interpretation of the Multinomial Logit

- By default, `nnet` and `mlogit` use the lowest category (i.e. 1 in an variables of 1, 2, 3) as the baseline
- This is rather atheoretical, so we should always select the most theoretically appropriate category for purposes of comparison (if possible)
  - We use the `relevel()` function
- Moreover, to calculate predictions for the baseline category (regardless of which one is chosen) it must be estimated separately (using a different baseline)
- Remember that changing the baseline category necessarily changes the coefficients of the model (*When will they not change?*)

Interpretation of the Multinomial Logit

- By default, `nnet` and `mlogit` use the lowest category (i.e. 1 in an variables of 1, 2, 3) as the baseline
- This is rather atheoretical, so we should always select the most theoretically appropriate category for purposes of comparison (if possible)
  - We use the `relevel()` function
- Moreover, to calculate predictions for the baseline category (regardless of which one is chosen) it must be estimated separately (using a different baseline)
- Remember that changing the baseline category necessarily changes the coefficients of the model (*When will they not change?*)
- Most analysts simply exclude a discussion of the baseline category (often requires a theoretical reason to justify picking one category as the baseline)

Estimating a Multinomial Logit in R

- There are two primary ways to estimate this in R:
  - **mlogit** package
  - **nnet** package
- **mlogit** requires a great deal more effort in data cleaning and preprocessing
- **nnet** estimates converge to **mlogit**

Estimating a Multinomial Logit in R

- There are two primary ways to estimate this in R:
  - **mlogit** package
  - **nnet** package
- **mlogit** requires a great deal more effort in data cleaning and preprocessing
- **nnet** estimates converge to **mlogit**
- Both are covered in the R tutorial – only **nnet** is discussed here

Estimating a Multinomial Logit in R

- Best practice is to always specify a new variable in your data frame to set the baseline category using the `relevel` function
  - `df$new_outcome <- relevel(df$outcome, ref = "Outcome Category")`

Estimating a Multinomial Logit in R

- Using the releveled outcome variable, can then run your multinomial logistic regression using `multinom()`
    - `multinom(formula, data, ..., Hess, censored, ...)`
    - Mostly standard options, only exception being `Hess = TRUE/FALSE` which you'll need to specify to TRUE

Estimating a Multinomial Logit in R

- `multinom(releveled_outcome ~ IV1 + IV2 + ..., data=df, Hess=TRUE)`

Estimating a Multinomial Logit in R

- `multinom(releveled_outcome ~ IV1 + IV2 + ..., data=df, Hess=TRUE)`
- Outcome here is a three category vote choice in the 1997 British Election – Liberal Democrats, Labour Party, Conservative Party
  - Baseline is Liberal Democrats in the following example

```
Call:
multinom(formula = voteD ~ gender + age + economic.cond.national +
    economic.cond.household, data = beps, Hess = TRUE)

Coefficients:
             (Intercept)       gender          age economic.cond.national economic.cond.household
Conservative   0.9895601 -0.088542495  0.016877636            -0.4461977             -0.04890819
Labour        -1.4964031 -0.004513461 -0.001296854             0.4633300              0.23645912

Residual Deviance: 2981.004
AIC: 3001.004
```

Estimating a Multinomial Logit in R

```
Call:
multinom(formula = voteD ~ gender + age + economic.cond.national +
    economic.cond.household, data = beps, Hess = TRUE)

Coefficients:
             (Intercept)       gender          age economic.cond.national economic.cond.household
Conservative   0.9895601 -0.088542495  0.016877636             -0.4461977             -0.04890819
Labour        -1.4964031 -0.004513461 -0.001296854              0.4633300              0.23645912

Residual Deviance: 2981.004
AIC: 3001.004
```

- Note that there are two sets of coefficients in these models
- These are **in comparison to the baseline category**

Estimating a Multinomial Logit in R

```
Call:
multinom(formula = voteD ~ gender + age + economic.cond.national +
    economic.cond.household, data = beps, Hess = TRUE)

Coefficients:
             (Intercept)       gender          age economic.cond.national economic.cond.household
Conservative   0.9895601 -0.088542495  0.016877636             -0.4461977             -0.04890819
Labour        -1.4964031 -0.004513461 -0.001296854              0.4633300              0.23645912

Residual Deviance: 2981.004
AIC: 3001.004
```

- Note that there are two sets of coefficients in these models
- These are **in comparison to the baseline category**
- **nnet** does not provide *p*-values after estimation

Estimating a Multinomial Logit in R

- To find statistical significance:
  - z_score <- summary(multinom_object)$coefficients / summary(multinom_object)$standard.errors
  - p_value <- (1 - pnorm(abs(z_score), 0, 1)) * 2

Estimating a Multinomial Logit in R

- To find statistical significance:
  - z_score < − summary(multinom_object)$coefficients / summary(multinom_object)$standard.errors
  - p_value < − (1 - pnorm(abs(z_score), 0, 1)) * 2
  - This output will take the form of a matrix
  - This isn't necessary for stargazer, as the function will calculate significance for you

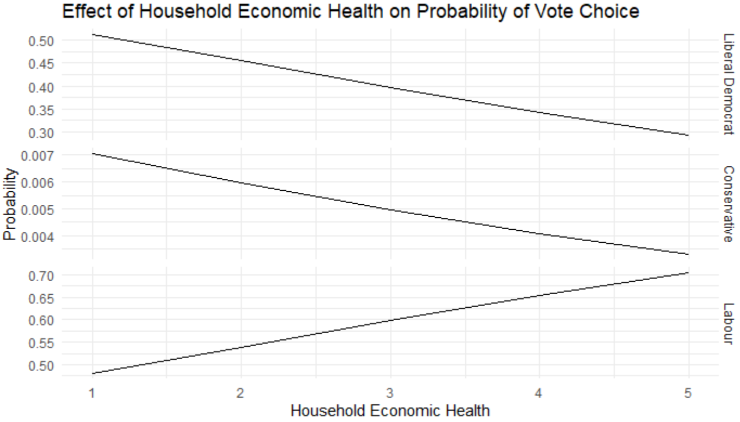
Multinomial Logit Results

Table: Effect of Perception of Economic Conditions, Gender, and Age on Vote Choice in 1997 British Elections

Baseline Category:	Liberal Democrat		Labour		Conservative	
	Conservative	Labour	Conservative	Liberal Democrat	Labour	Liberal Democrat
Male	−0.089	−0.005	−0.084	0.005	0.084	0.089
	(0.146)	(0.134)	(0.129)	(0.134)	(0.129)	(0.146)
Age	0.017***	−0.001	0.018***	0.001	−0.018***	−0.017***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Perception of National Economic Health	−0.446***	0.463***	−0.910***	−0.463***	0.910***	0.446***
	(0.090)	(0.086)	(0.083)	(0.086)	(0.083)	(0.090)
Perception of Household Economic Health	−0.049	0.236***	−0.285***	−0.236***	0.285***	0.049
	(0.083)	(0.077)	(0.075)	(0.077)	(0.075)	(0.083)
Constant	0.990**	−1.496***	2.486***	1.497***	−2.486***	−0.990**
	(0.451)	(0.434)	(0.413)	(0.434)	(0.413)	(0.451)
N	1525	1525	1525	1525	1525	1525
AIC	3,001.004	3,001.004	3,001.004	3,001.004	3,001.004	3,001.004
Note:						*p<0.1; **p<0.05; ***p<0.01

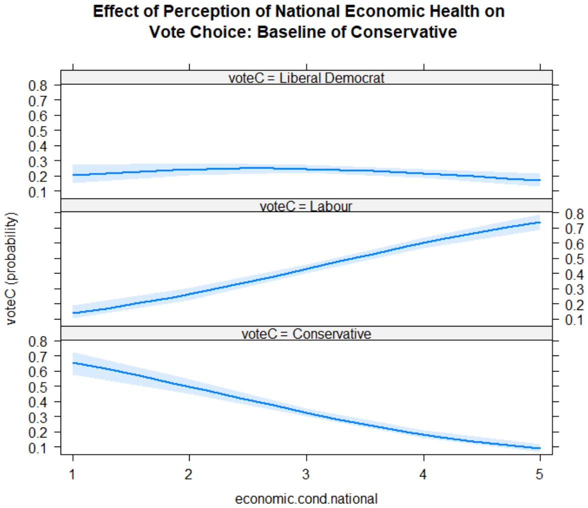
Estimating a Multinomial Logit in R

- Can use the predicted probabilities procedure from last week to create predicted probability figures



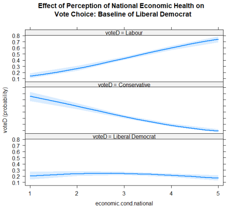
Estimating a Multinomial Logit in R

- Can use effects package to create the same graph with confidence intervals

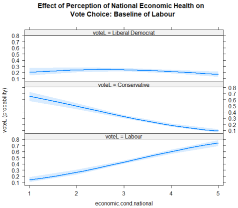


Estimating a Multinomial Logit in R

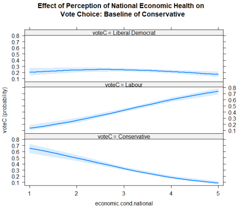
- If correctly specified, the baseline category should only affect the appearance, not the substance of predicted probabilities figures



(a) Liberal Democracy



(b) Labour



(c) Conservative

Figure: Predicted Probabilities by Different Baseline Categories

Estimating a Multinomial Logit in R

- Can also use the results from the `Effect()` command to create ggplots with confidence intervals
- Requires a good deal of understanding in ggplot, but is possible

Independence of Irrelevant Alternatives

- Calculating a multinomial logit requires making the independence of irrelevant alternatives assumption

Independence of Irrelevant Alternatives

- Calculating a multinomial logit requires making the independence of irrelevant alternatives assumption
- For illustration, consider that we divide one of the original 3 categories from our party example into two separate categories

Independence of Irrelevant Alternatives

- Calculating a multinomial logit requires making the independence of irrelevant alternatives assumption
- For illustration, consider that we divide one of the original 3 categories from our party example into two separate categories
- Such that we have a 4 category dependent variable:
    1. (A) Acción Democrática
    2. (B) Bolivarian Movement
    3. (C) COPEI
    4. (D) Democrático Party

Independence of Irrelevant Alternatives

- Can we simply assume that the probabilities of choosing an alternative party remain consistent from the earlier calculation?

Independence of Irrelevant Alternatives

- Can we simply assume that the probabilities of choosing an alternative party remain consistent from the earlier calculation?
- More formally, can we assume *ALL* of the following:
  1. $\Pr A$ is unchanged
  2. $\Pr B = \Pr(Bol.Mov.) + \Pr(Dem.)$
  3. $\Pr C$ is unchanged

Independence of Irrelevant Alternatives

- Can we simply assume that the probabilities of choosing an alternative party remain consistent from the earlier calculation?
- More formally, can we assume *ALL* of the following:
  1. $\Pr A$ is unchanged
  2. $\Pr B = \Pr(Bol.Mov.) + \Pr(Dem.)$
  3. $\Pr C$ is unchanged
- The sample of observations remains same with $\text{N}_B = \text{N}_{(B)} + \text{N}_{(D)}$

Independence of Irrelevant Alternatives

- However, our potential problem (and the IIA assumption) has nothing to do with the sample of observations, but rather with the characteristics in choosing alternatives

Independence of Irrelevant Alternatives

- However, our potential problem (and the IIA assumption) has nothing to do with the sample of observations, but rather with the characteristics in choosing alternatives
- The IIA assumption involves potential correlation of the error terms (which are themselves assumed to be non-correlated)

Independence of Irrelevant Alternatives

- However, our potential problem (and the IIA assumption) has nothing to do with the sample of observations, but rather with the characteristics in choosing alternatives
- The IIA assumption involves potential correlation of the error terms (which are themselves assumed to be non-correlated)
- If IIA is violated, then the errors are correlated

Independence of Irrelevant Alternatives

- However, our potential problem (and the IIA assumption) has nothing to do with the sample of observations, but rather with the characteristics in choosing alternatives
- The IIA assumption involves potential correlation of the error terms (which are themselves assumed to be non-correlated)
- If IIA is violated, then the errors are correlated
- This leads to inconsistent estimates

Independence of Irrelevant Alternatives

- To illustrate, let us define the probability of voting for Acción Democrática (A) before the introduction of the new alternative:

$$\Pr A = \frac{e^{\mathbf{X}\beta_A}}{1 + e^{\mathbf{X}\beta_A} + e^{\mathbf{X}\beta_B}}$$

- If we include a new alternative, and if that alternative is irrelevant, then we simply add a new category (not a problem)

$$\Pr A = \frac{e^{\mathbf{X}\beta_A}}{1 + e^{\mathbf{X}\beta_A} + e^{\mathbf{X}\beta_{(Bol.Mov.)}} + e^{\mathbf{X}\beta_{(Dem.)}}}$$

Independence of Irrelevant Alternatives

- However, if the alternative theoretically should not have an impact, but in reality does because $\beta_B = \beta_{Bol.Mov.} = \beta_{Dem.}$ then we have a problem because the new probabilities become:
$$\Pr A = \frac{e^{\mathbf{X}\beta_A}}{1 + e^{\mathbf{X}\beta_A} + 2e^{\mathbf{X}\beta_B}}$$

Independence of Irrelevant Alternatives

- However, if the alternative theoretically should not have an impact, but in reality does because $\beta_B = \beta_{Bol.Mov.} = \beta_{Dem.}$ then we have a problem because the new probabilities become:

$$\Pr A = \frac{e^{\mathbf{X}\beta_A}}{1 + e^{\mathbf{X}\beta_A} + 2e^{\mathbf{X}\beta_B}}$$

- For many (possibly most) political phenomena, adding a new alternative often causes problems for our inferences if we rely on multinomial logit

Independence of Irrelevant Alternatives

- However, if the alternative theoretically should not have an impact, but in reality does because $\beta_B = \beta_{Bol.Mov.} = \beta_{Dem.}$ then we have a problem because the new probabilities become:

$$\Pr A = \frac{e^{\mathbf{X}\beta_A}}{1 + e^{\mathbf{X}\beta_A} + 2e^{\mathbf{X}\beta_B}}$$

- For many (possibly most) political phenomena, adding a new alternative often causes problems for our inferences if we rely on multinomial logit
- Therefore we need to find alternative methods of estimation

Independence of Irrelevant Alternatives

- Determining whether a violation of the IIA assumption has occurred, essentially involves testing whether two outcomes (alternatives) can be combined

Independence of Irrelevant Alternatives

- Determining whether a violation of the IIA assumption has occurred, essentially involves testing whether two outcomes (alternatives) can be combined
- If category m is indistinguishable from category n (i.e. the Bolivarian Movement and Democrático Party), then we can test whether the coefficients are equal

Independence of Irrelevant Alternatives

- Determining whether a violation of the IIA assumption has occurred, essentially involves testing whether two outcomes (alternatives) can be combined
- If category m is indistinguishable from category n (i.e. the Bolivarian Movement and Democrático Party), then we can test whether the coefficients are equal
- Formally, we test the following null hypothesis:
- $H_0 : \beta_m = \beta_n$ or $\beta_m - \beta_n = 0$ or $(\beta_{1,m|j} - \beta_{1,n|j}) = 0$
  - where j is the baseline category

Testing for IIA Violations

- Hausman Test
    - Run fully specified model (including all categories minus a baseline) and save results

Testing for IIA Violations

- Hausman Test
  - Run fully specified model (including all categories minus a baseline) and save results
  - Run second model that eliminates one category and calculate Hausman statistic

Testing for IIA Violations

- Hausman Test
  - Run fully specified model (including all categories minus a baseline) and save results
  - Run second model that eliminates one category and calculate Hausman statistic
  - Recall that Hausman test is distributed χ^2 and is calculated using the following:
  - $\text{H} = (\beta_C - \beta_E)'(V_C - V_E)^{-1}(\beta_C - \beta_E)$

Testing for IIA Violations

- Hausman Test
  - Run fully specified model (including all categories minus a baseline) and save results
  - Run second model that eliminates one category and calculate Hausman statistic
  - Recall that Hausman test is distributed χ^2 and is calculated using the following:
  - $\text{H} = (\beta_C - \beta_E)'(V_C - V_E)^{-1}(\beta_C - \beta_E)$
  - In R this is done with `hmftest()` after `mlogit()` `assignment`
    - Specify an unconstrained and constained mlogit object
    - hmftest(unconstrained, constrained)

Alternatives to Multinomial Logit — Multinomial Probit

- This model is more resistant to violations of the IIA assumption because the disturbances are distributed according to the multivariate normal distribution

Alternatives to Multinomial Logit — Multinomial Probit

- This model is more resistant to violations of the IIA assumption because the disturbances are distributed according to the multivariate normal distribution
- This allows for categories of the dependent variable to vary without affecting the remaining coefficients

Alternatives to Multinomial Logit — Multinomial Probit

- This model is more resistant to violations of the IIA assumption because the disturbances are distributed according to the multivariate normal distribution
- This allows for categories of the dependent variable to vary without affecting the remaining coefficients
- Mathematically, this model is extremely more difficult to compute (even for computers)

Alternatives to Multinomial Logit — Multinomial Probit

- This model is more resistant to violations of the IIA assumption because the disturbances are distributed according to the multivariate normal distribution
- This allows for categories of the dependent variable to vary without affecting the remaining coefficients
- Mathematically, this model is extremely more difficult to compute (even for computers)
  - Involves calculating an integral for each category comparison

Alternatives to Multinomial Logit — Multinomial Probit

- This model is more resistant to violations of the IIA assumption because the disturbances are distributed according to the multivariate normal distribution
- This allows for categories of the dependent variable to vary without affecting the remaining coefficients
- Mathematically, this model is extremely more difficult to compute (even for computers)
  - Involves calculating an integral for each category comparison
  - Becomes extremely cumbersome with numerous (more than 3 or 4) categories in the dependent variable

Alternatives to Multinomial Logit — Multinomial Probit

- This model is more resistant to violations of the IIA assumption because the disturbances are distributed according to the multivariate normal distribution
- This allows for categories of the dependent variable to vary without affecting the remaining coefficients
- Mathematically, this model is extremely more difficult to compute (even for computers)
  - Involves calculating an integral for each category comparison
  - Becomes extremely cumbersome with numerous (more than 3 or 4) categories in the dependent variable
- In the **mlogit** package: `mlogit(formula, data, ...` **probit=TRUE**)
  - Discussed in the tutorial

Alternatives to Multinomial Logit — Conditional Logit

- Sometimes called Conditional Fixed Effects Logit

Alternatives to Multinomial Logit — Conditional Logit

- Sometimes called Conditional Fixed Effects Logit
- Allows individual (or group) specific effects (i.e. fixed effects)

Alternatives to Multinomial Logit — Conditional Logit

- Sometimes called Conditional Fixed Effects Logit
- Allows individual (or group) specific effects (i.e. fixed effects)
  - This avoids the IIA assumption by linking calculations to specific groups

Alternatives to Multinomial Logit — Conditional Logit

- Sometimes called Conditional Fixed Effects Logit
- Allows individual (or group) specific effects (i.e. fixed effects)
  - This avoids the IIA assumption by linking calculations to specific groups
  - Therefore, the characteristics of the alternatives are linked directly to specific individuals (groups)

Alternatives to Multinomial Logit — Conditional Logit

- Sometimes called Conditional Fixed Effects Logit
- Allows individual (or group) specific effects (i.e. fixed effects)
  - This avoids the IIA assumption by linking calculations to specific groups
  - Therefore, the characteristics of the alternatives are linked directly to specific individuals (groups)
- Essentially, this model computes a multinomial calculation for every observed group

Alternatives to Multinomial Logit — Conditional Logit

- Sometimes called Conditional Fixed Effects Logit
- Allows individual (or group) specific effects (i.e. fixed effects)
  - This avoids the IIA assumption by linking calculations to specific groups
  - Therefore, the characteristics of the alternatives are linked directly to specific individuals (groups)
- Essentially, this model computes a multinomial calculation for every observed group
  - Multinomial logit has $J-1$ parameters β_{km} for each $\mathbf{X}_k$ but only a single value of $\mathbf{X}_k$ for each individual

Alternatives to Multinomial Logit — Conditional Logit

- Sometimes called Conditional Fixed Effects Logit
- Allows individual (or group) specific effects (i.e. fixed effects)
  - This avoids the IIA assumption by linking calculations to specific groups
  - Therefore, the characteristics of the alternatives are linked directly to specific individuals (groups)
- Essentially, this model computes a multinomial calculation for every observed group
  - Multinomial logit has $J-1$ parameters β_{km} for each $\mathbf{X}_k$ but only a single value of $\mathbf{X}_k$ for each individual
  - Conditional logit has a single β_k for each variable $\mathbf{X}_k$ but there are J values of the variable for each individual

Alternatives to Multinomial Logit — Conditional Logit

- Sometimes called Conditional Fixed Effects Logit
- Allows individual (or group) specific effects (i.e. fixed effects)
  - This avoids the IIA assumption by linking calculations to specific groups
  - Therefore, the characteristics of the alternatives are linked directly to specific individuals (groups)
- Essentially, this model computes a multinomial calculation for every observed group
  - Multinomial logit has $J-1$ parameters β_{km} for each $\mathbf{X}_k$ but only a single value of $\mathbf{X}_k$ for each individual
  - Conditional logit has a single β_k for each variable $\mathbf{X}_k$ but there are J values of the variable for each individual
  - Example: estimating Supreme Court behavior across 9 justices

Alternatives to Multinomial Logit — Conditional Logit

- Sometimes called Conditional Fixed Effects Logit
- Allows individual (or group) specific effects (i.e. fixed effects)
  - This avoids the IIA assumption by linking calculations to specific groups
  - Therefore, the characteristics of the alternatives are linked directly to specific individuals (groups)
- Essentially, this model computes a multinomial calculation for every observed group
  - Multinomial logit has $J-1$ parameters β_{km} for each $\mathbf{X}_k$ but only a single value of $\mathbf{X}_k$ for each individual
  - Conditional logit has a single β_k for each variable $\mathbf{X}_k$ but there are J values of the variable for each individual
  - Example: estimating Supreme Court behavior across 9 justices
- **mclogit** package