Multinomial and ordered logits

Silje Synnøve Lyder Hermansen

16-03-2023

GLM: A recap

Reminder: What is a GLM?

Regressions aim to describe (a linear) relationship between x and y with one number, β.

Reminder: What is a GLM?

Regressions aim to describe (a linear) relationship between x and y with one number, β.

- Assumes a continuous and unbounded variable.

Reminder: What is a GLM?

Regressions aim to describe (a linear) relationship between x and y with one number, β.

- Assumes a continuous and unbounded variable.
- When y is neither (e.g. binary), we relied on a latent continuous variable

Reminder: What is a GLM?

Regressions aim to describe (a linear) relationship between x and y with one number, β.

- Assumes a continuous and unbounded variable.
- When y is neither (e.g. binary), we relied on a latent continuous variable
- To approximate the latent variable, we calculated the logodds (i.e. we compare)

Reminder: What is a GLM?

Regressions aim to describe (a linear) relationship between x and y with one number, β.

- Assumes a continuous and unbounded variable.
- When y is neither (e.g. binary), we relied on a latent continuous variable
- To approximate the latent variable, we calculated the logodds (i.e. we compare)

$\Rightarrow$ *Probability distribution maps unobserved variable to observed outcomes.*

Ordered logistic regression

What is an ordered variable?

A ranked variable with unknown distance between categories.

What is an ordered variable?

A ranked variable with unknown distance between categories.

- Often the result of binning: Close connection to latent formulation.

What is an ordered variable?

A ranked variable with unknown distance between categories.

- Often the result of binning: Close connection to latent formulation.
- We can choose how to treat it: As linear, categorical or **ordinal**.

What is an ordered variable?

A ranked variable with unknown distance between categories.

- Often the result of binning: Close connection to latent formulation.
- We can choose how to treat it: As linear, categorical or **ordinal**.

$\Rightarrow$ *estimate a single set of regression parameters, but keep the information on the order without assuming a continuous variable.*

Two conceptions of ordered logisitc regression

There are two ways of uncerstanding the ordered logit:

Two conceptions of ordered logisitc regression

There are two ways of uncerstanding the ordered logit:

- Latent variable: useful for interpretation.

Two conceptions of ordered logisitc regression

There are two ways of uncerstanding the ordered logit:

- Latent variable: useful for interpretation.
- Parallel regressions: useful for understanding and checking estimation.

Two conceptions of ordered logisitc regression

There are two ways of uncerstanding the ordered logit:

- Latent variable: useful for interpretation.
- Parallel regressions: useful for understanding and checking estimation.

Latent variable approach: cutpoints

Cutpoints

We rely on cutpoints to slice up the latent variable and determine outcomes

Cutpoints

We rely on cutpoints to slice up the latent variable and determine outcomes

- **Binomial logistic:** One cutpoint. $\rightarrow$ Rarely estimated.

Cutpoints

We rely on cutpoints to slice up the latent variable and determine outcomes

- **Binomial logistic:** One cutpoint. $\rightarrow$ Rarely estimated.
- **Ordinal logistic:** Serveral cutpoints. $\rightarrow$ Explicit.

Cutpoints

We rely on cutpoints to slice up the latent variable and determine outcomes

- **Binomial logistic:** One cutpoint. $\rightarrow$ Rarely estimated.
- **Ordinal logistic:** Serveral cutpoints. $\rightarrow$ Explicit.

$\Rightarrow$ *Model estimates both regression parameters (β) and cutpoints (τ).*

A series of cutpoints

You are in the category m **when the latent variable is between its two cutpoints:** $\tau_{m-1} < y^{\star} < \tau_m$

A series of cutpoints

You are in the category m **when the latent variable is between its two cutpoints:** $\tau_{m-1} < y^{\star} < \tau_m$

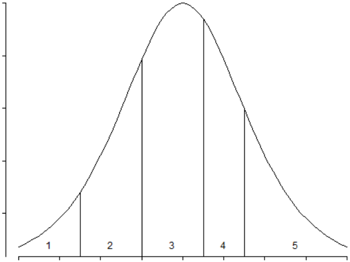


Figure 1: Slicing up a latent variable

The regression coefficients

The model calculates the odds of being lower than τ_m

The regression coefficients

The model calculates the odds of being lower than τ_m

- The first cutpoint (τ_0) is 0 ($-inf$): you cant be lower than the lowest.

The regression coefficients

The model calculates the odds of being lower than τ_m

- The first cutpoint (τ_0) is 0 ($-inf$): you cant be lower than the lowest.
- The last cutpoint is 1 ($+inf$): all observations are in some category.

The regression coefficients

The model calculates the odds of being lower than τ_m

- The first cutpoint (τ_0) is 0 ($-inf$): you cant be lower than the lowest.
- The last cutpoint is 1 ($+inf$): all observations are in some category.
- You end up with $m-1$ cutpoints.

The regression output

The regression output reports both β and τ

The regression output

The regression output reports both β and τ

- **Regression coefficient** β is reported in relation to *upper* cutpoint of the category: $\tau_m - \beta x_i$

The regression output

The regression output reports both β and τ

- **Regression coefficient** β is reported in relation to *upper* cutpoint of the category: $\tau_m - \beta x_i$
- **Cutpoints** serve also as intercepts.

The regression output

The regression output reports both β and τ

- **Regression coefficient** β is reported in relation to *upper* cutpoint of the category: $\tau_m - \beta x_i$
- **Cutpoints** serve also as intercepts.

The predicted value

The predicted probability of being in category *m*:

The predicted value

The predicted probability of being in category m:

$$Pr(y_i = m) = \frac{exp(\tau_m - \beta x_i)}{1 + exp(\tau_m - \beta x_i)} - \frac{exp(\tau_{m-1} - \beta x_i)}{1 + exp(\tau_{m-1} - \beta x_i)} \qquad (1)$$

An example: Attitudes towards redistribution

An example:

ESS respondents (that voted V or DF) are asked to what extent they believe the state should engage in redistribution (1 = disagree; 5 = agree).

```
#Load in data
download.file(
  url("https://siljehermansen.github.io/teaching/beyond-linear-models/kap10
  destfile = "kap10.rda"
)
df <- kap10

#Check distribution
barplot(table(df$Udjaevn))
```

An example:

ESS respondents (that voted V or DF) are asked to what extent they believe the state should engage in redistribution (1 = disagree; 5 = agree).

An example:

ESS respondents (that voted V or DF) are asked to what extent they believe the state should engage in redistribution (1 = disagree; 5 = agree).

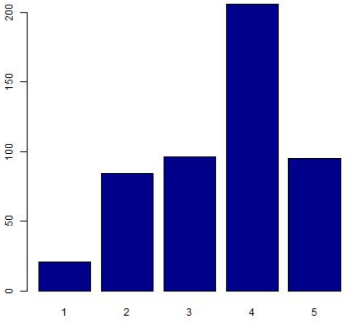


Figure 2: Attitudes towards redistribution is an ordered variable

Attitudes towards redistribution as a function of income

```
#Library for ordinal regression
library(MASS)
#Recode into ordered factor
df$Udjaevn.ord <- as.ordered(as.factor(df$Udjaevn))
#Run regression
mod.ord <- polr(Udjaevn.ord ~ Indtaegt,
                df,
                method = "logistic",
                Hess = TRUE)
summary(mod.ord)
```

Attitudes towards redistribution as a function of income

```
## Call:
## polr(formula = Udjaevn.ord ~ Indtaegt, data = df, Hess = TRUE,
##     method = "logistic")
##
## Coefficients:
##           Value Std. Error t value
## Indtaegt 0.1153    0.03155   3.653
##
## Intercepts:
##     Value   Std. Error t value
## 1|2 -2.4186  0.2903    -8.3306
## 2|3 -0.6008  0.2179    -2.7566
## 3|4  0.3069  0.2150     1.4277
## 4|5  2.2276  0.2403     9.2686
##
## Residual Deviance: 1298.396
## AIC: 1308.396
## (51 observations deleted due to missingness)
```

We learn two things from the regression output

Regression coefficient reports effect of x on probability to be placed one category higher

We learn two things from the regression output

Regression coefficient reports effect of x on probability to be placed one category higher

- Effect in logodds: `0.115`

We learn two things from the regression output

Regression coefficient reports effect of x on probability to be placed one category higher

- Effect in logodds: `0.115`
- We can backtransform to one unit increase in x: $(exp(\beta) - 1) \times 100$ $= 12\%$ increase in likelihood of a higher category.

We learn two things from the regression output

Regression coefficient reports effect of x on probability to be placed one category higher

- Effect in logodds: `0.115`
- We can backtransform to one unit increase in x: $(exp(\beta) - 1) \times 100$ $= 12\%$ increase in likelihood of a higher category.

$\Rightarrow$ *Hypothesis testing as in a binomial logit*

We learn two things from the regression output

We have one intercept per cutpoint

We learn two things from the regression output

We have one intercept per cutpoint

- e.g.: intercept of passing from 1 to 2 is `-2.419`

We learn two things from the regression output

We have one intercept per cutpoint

- e.g.: intercept of passing from 1 to 2 is `-2.419`
- e.g.: intercept is reported as significant (with standard errors)

We learn two things from the regression output

We have one intercept per cutpoint

- e.g.: intercept of passing from 1 to 2 is `-2.419`
- e.g.: intercept is reported as significant (with standard errors)

$\Rightarrow$ *The model does a fair job in distinguishing between categories.*

Predicted scenarios

We interpret predicted probability by choosing one level of x and one category (two cutpoints) of y: What is the probability of m?

$$Pr(y_i = m) = \frac{exp(\tau_m - \beta x_i)}{1 + exp(\tau_m - \beta x_i)} - \frac{exp(\tau_{m-1} - \beta x_i)}{1 + exp(\tau_{m-1} - \beta x_i)} \qquad (2)$$

Example

Let's choose low-income respondents (x = 1) and category 3 (diff between cutpoints 2 and 3)

```
z = mod.ord$zeta
x = 1

logodds1 <- z[3] - coefficients(mod.ord) * x
logodds2 <- z[3-1] - coefficients(mod.ord) * x
## Probabilities
p1 <- exp(logodds1)/(1 + exp(logodds1)) #3/4 or lower
p2 <- exp(logodds2)/(1 + exp(logodds2)) #2/3 or lower
## Difference between cutpoints
p1 - p2 #cat 3
```

An example

Predicted proportion in category

```
paste(round((p1-p2)*100),
"% of low-income respondents are predicted to answer x = 3 ('neutral')." )
```

[1] "22 % of low-income respondents are predicted to answer x = 3 ('neutral')."

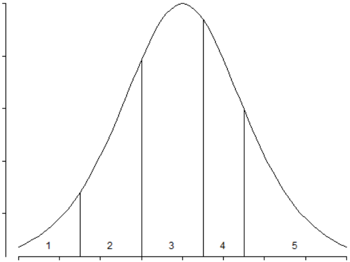
Cumulative probability

```
paste(round((p1)*100),
"% of low-income respondents are predicted to answer x = 3 ('neutral') or l
```

[1] "55 % of low-income respondents are predicted to answer x = 3 ('neutral') or lower to the question of whether they support redistribution."

Two ways of viewing the slicing

We can report the probability (e.g. 0.22) of ending up between two cutpoints, or the *cumulative* probability (e.g. 0.55) to be below each



point.

Exercice:

Increase the τ (z) within each value of Income (x)

```
##Create empty plot
plot(y = 0,
     x = 0,
     axes = FALSE,
     xlim = c(1,4),
     ylim = c(0,1),
     ylab = "Probability of z or below",
     xlab = "Thresholds",
     main = "Cumulative probability \nof support for redistribution",
     type = "n")
axis(1, at = 1:length(p1),
     labels = names(p1))
axis(2)
```

Exercice:

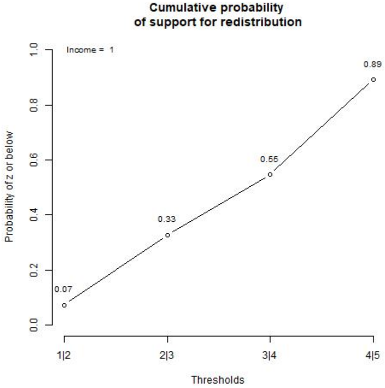
Increase the τ (z) within each value of Income (x)

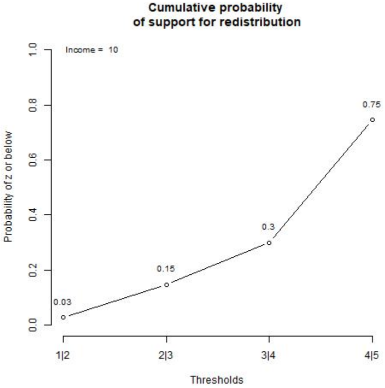
```
#Set values for prediction
x = 10 #Let this go from 1 to 10; check the shape of 10
z = mod.ord$zeta
#Logodds
logodds1 <- z - coefficients(mod.ord) * x
#Probabilities
p1 <- exp(logodds1)/(1 + exp(logodds1)) #3/4 or lower

#Plot probabilities
lines(y = p1,
      x = 1:length(p1),
      type = "b")
#Set legend (report x-value)
legend("topleft",
       bty = "n",
       cex = 0.8,
     paste("Income = ", x))

#Plot probabilities
text(y = p1+0.06,
```

Result


Cumulative probability
of support for redistribution
Income = 1
Probability of z or below
0.0
0.2
0.4
0.6
0.8
1.0
0.07
0.33
0.55
0.89
1|2
2|3
3|4
4|5
Thresholds


Cumulative probability
of support for redistribution
Income = 10
Probability of z or below
0.0
0.2
0.4
0.6
0.8
1.0
0.03
0.15
0.3
0.75
1|2
2|3
3|4
4|5
Thresholds

Parallel regressions approach: for assessment

Parallel regressions approach

The parallel regression approach is useful to understand how the model is estimated

Parallel regressions approach

The parallel regression approach is useful to understand how the model is estimated

- The y is recoded into $m-1$ dummy variables indicating if $y \leq m$

Parallel regressions approach

The parallel regression approach is useful to understand how the model is estimated

- The y is recoded into $m-1$ dummy variables indicating if $y \leq m$
- Run a series of regressions where all β are fixed (i.e.: the same).

Parallel regressions approach

The parallel regression approach is useful to understand how the model is estimated

- The y is recoded into $m-1$ dummy variables indicating if $y \leq m$
- Run a series of regressions where all β are fixed (i.e.: the same).

$\Rightarrow$ *This is also useful when we assess the model*

How good is our model?

The basic assumption

The basic assumption is that all parallel regressions have (about) the same regression coefficient

The basic assumption

The basic assumption is that all parallel regressions have (about) the same regression coefficient

- Check the mean of the predictor for each value of y. Does it trend?

The basic assumption

The basic assumption is that all parallel regressions have (about) the same regression coefficient

- Check the mean of the predictor for each value of y. Does it trend?

```
df %>%
  filter(!is.na(Udjaevn)) %>%
  group_by(Udjaevn) %>%
  summarize(mean(Indtaegt, na.rm = T))
```

```
## # A tibble: 5 x 2
##   Udjaevn `mean(Indtaegt, na.rm = T)`
##     <dbl>                       <dbl>
## 1       1                        4.8
## 2       2                        5.58
## 3       3                        5.96
## 4       4                        6.41
## 5       5                        6.75
```

The basic assumption

The basic assumption is that all parallel regressions have (about) the same regression coefficient

- Check the mean of the predictor for each value of y. Does it trend?

```
df %>%
  filter(!is.na(Udjaevn)) %>%
  group_by(Udjaevn) %>%
  summarize(mean(Indtaegt, na.rm = T))
```

```
## # A tibble: 5 x 2
##   Udjaevn `mean(Indtaegt, na.rm = T)`
##     <dbl>                       <dbl>
## 1       1                        4.8
## 2       2                        5.58
## 3       3                        5.96
## 4       4                        6.41
## 5       5                        6.75
```

- Run parallel regressions without contstraint on β. Are they similar?

An example of parallel regressions

Recode into dummies

The dummies flag cases below a cumulative threshold of *outcomes*

```
##
df$ut1 <- ifelse(df$Udjaevn > 1, 1 , 0) #2 or above
df$ut2 <- ifelse(df$Udjaevn > 2, 1 , 0) #3 or above
df$ut3 <- ifelse(df$Udjaevn > 3, 1 , 0) #4 or above
df$ut4 <- ifelse(df$Udjaevn > 4, 1 , 0) #5
```

$\Rightarrow$ *The model then runs 4 regressions where* β *reports an aggregated value from all 4 coefficients (think: weigted mean).*

Run four regressions

Let's examplify with the parallel regressions without fixed β:

```
##Parallel regressions:
mod1 <- glm(ut1 ~ Indtaegt, df, family = "binomial")
mod2 <- glm(ut2 ~ Indtaegt, df, family = "binomial")
mod3 <- glm(ut3 ~ Indtaegt, df, family = "binomial")
mod4 <- glm(ut4 ~ Indtaegt, df, family = "binomial")
```

Compare coefficients from four regressions

```
##
## =====================================================
##                       Dependent variable:
##                 -------------------------------------
##                    ut1      ut2      ut3       ut4
##                    (1)      (2)      (3)       (4)
## -----------------------------------------------------
## Indtaegt         0.189**  0.125*** 0.110***  0.094**
##                  (0.085)  (0.041)  (0.035)   (0.045)
##
## Constant         2.048*** 0.552**   -0.270  -2.082***
##                  (0.474)  (0.260)  (0.231)   (0.319)
##
## -----------------------------------------------------
## Observations       459      459      459       459
## Log Likelihood   -79.653  -234.669 -303.983 -217.674
## Akaike Inf. Crit. 163.306  473.338  611.967   439.348
## =====================================================
## Note:                     *p<0.1; **p<0.05; ***p<0.01
```

Coefficient should be a weighted average from four regressions

These βs are weighted by the number of observations in each category:

```
table(df$Udjaevn)
```

```
##
##   1   2   3   4   5
##  21  84  96 206  95
```

We can plot the βs for comparison:

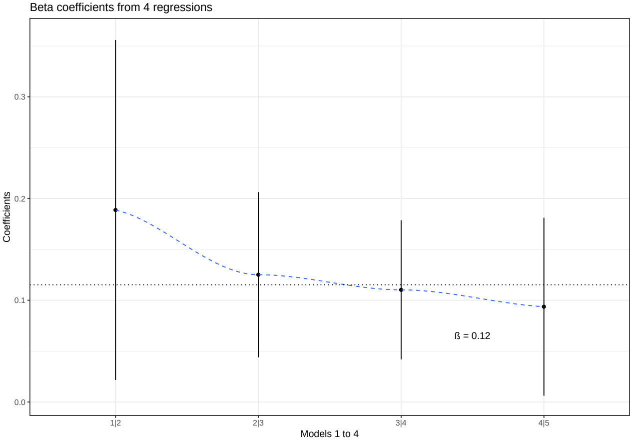
```
results <- rbind(summary(mod1)$coefficients[2, c(1,2)],
                 summary(mod2)$coefficients[2, c(1,2)],
                 summary(mod3)$coefficients[2, c(1,2)],
                 summary(mod4)$coefficients[2, c(1,2)])
thresholds <- c("1|2","2|3","3|4","4|5")
```

We can plot the βs for comparison:

```
ggplot() +
  geom_point(aes(y = results[, "Estimate"],
               x = thresholds)) +
  geom_smooth(aes(y = results[, "Estimate"],
                x = 1:4),
              lty = 2,
              lwd = 0.5) +
  geom_segment(aes(x = 1:4,
               xend = 1:4,
               y = results[, "Estimate"]-results[, "Std. Error"]*1.96,
               yend = results[, "Estimate"]+results[, "Std. Error"]*1.96)) +
  theme_bw() +
  ylim(c(results[, "Estimate"][4]-results[, "Std. Error"][4]*2,
         results[, "Estimate"][1]+results[, "Std. Error"][1]*2)) +
  geom_hline(yintercept = mod.ord$coefficients,
             lty = 3) +
  geom_text(aes(y = mod.ord$coefficients-0.05,
                x = 3.5,
                label = paste("\u03b2 =", round(mod.ord$coefficients,2))
                ),
            parse = F) +
  labs(title = "Beta coefficients from 4 regressions") +
  ylab("Coefficients") +
  xlab("Models 1 to 4")
```

We can plot the βs for comparison:

The overall β is `0.12`. If the ordered model describes the data well, then all the unconstrained βs should ressemble that description.



A visual inspection

A more visual way of checking the “parallel lines assumption” is to inspect if the regression lines are parallel.

When is it smart to run an ordered logit?

- You have few ordered categories

When is it smart to run an ordered logit?

- You have few ordered categories
- The effect is approximately the same across the categories (parallel lines assumption)

What do I do if the assumption doesn't hold?

- Run an OLS/linear model:
  - if you have many categories
  - fairly equal spread of observations between categories

What do I do if the assumption doesn't hold?

- Run an OLS/linear model:
  - if you have many categories
  - fairly equal spread of observations between categories
- Run a multinomial model:
  - i.e. estimate different β for each regression/threshold

Discrete choice models

Dependent variable: nominal

The discrete choice models describe mutually exclusive choices.

Dependent variable: nominal

The discrete choice models describe mutually exclusive choices.

- The choice variable is nominal: we cannot rank it

Dependent variable: nominal

The discrete choice models describe mutually exclusive choices.

- The choice variable is nominal: we cannot rank it
- Our *appreciation* of it is continuous. Two sets of models:

Dependent variable: nominal

The discrete choice models describe mutually exclusive choices.

- The choice variable is nominal: we cannot rank it
- Our *appreciation* of it is continuous. Two sets of models:
  - Multinomial: Models *chooser* characteristics

Dependent variable: nominal

The discrete choice models describe mutually exclusive choices.

- The choice variable is nominal: we cannot rank it
- Our *appreciation* of it is continuous. Two sets of models:
  - Multinomial: Models *chooser* characteristics
  - Conditional logit: Models *choice* characteristics

Multinomial logistic regression

Two conceptions of multinomial regression

Two conceptions of multinomial regression

Two conceptions of multinomial regression

- **A series of binomial logits** with the same reference category.
- **Latent variable approach:** Our utility of each choice.

Two conceptions of multinomial regression

A series of binomial logits with the same reference category.

Two conceptions of multinomial regression

A series of binomial logits with the same reference category.

- Data is subset to compare two groups $\rightarrow$ data/variation intensive model choice.

Two conceptions of multinomial regression

A series of binomial logits with the same reference category.

- Data is subset to compare two groups → data/variation intensive model choice.
- Categories/choice are mutually exclusive → Different β for each subset/choice

Two conceptions of multinomial regression

Two conceptions of multinomial regression

A series of binomial logits with the same reference category.

- Data is subset to compare two groups $\rightarrow$ data/variation intensive model choice.
- Categories/choice are mutually exclusive $\rightarrow$ Different β for each subset/choice

$\Rightarrow$ *All choices are given a probability and they sum up to one.*

Two conceptions of multinomial regression

Latent variable approach: Imagine k choices modeled as $y_m = \alpha_m \times \beta_m x$

- $\beta_m x_i$ reflects the utility of a choice k for the chooser i with x characteristic. $\rightarrow$ systematic term

Two conceptions of multinomial regression

Latent variable approach: Imagine k choices modeled as $y_m = \alpha_m \times \beta_m x$

- $\beta_m x_i$ reflects the utility of a choice k for the chooser i with x characteristic. $\rightarrow$ systematic term
- α_m reflects the baseline utility of that choice $\rightarrow$ stochastic term

Two conceptions of multinomial regression

Two conceptions of multinomial regression

Latent variable approach: Imagine k choices modeled as $y_m = \alpha_m \times \beta_m x$

- $\beta_m x_i$ reflects the utility of a choice k for the chooser i with x characteristic. $\rightarrow$ systematic term
- α_m reflects the baseline utility of that choice $\rightarrow$ stochastic term

$\Rightarrow$ *The preferred choice is the one with the highest utility because both or either are high*

Main assumption: IIA

Independence of irrelevant alternatives:

Main assumption: IIA

Independence of irrelevant alternatives:

- there are no choices beyond what is modeled

Main assumption: IIA

Independence of irrelevant alternatives:

- there are no choices beyond what is modeled
- consistency: if we prefer $A > B$ and $B > C$, then also $A > C$

Main assumption: IIA

Independence of irrelevant alternatives:

- there are no choices beyond what is modeled
- consistency: if we prefer A $>$ B and B $>$ C, then also A $>$ C

$\Rightarrow$ *The β does not depend on on other values of y (other alternatives).*

Testing the main assumption:

The Hausmann-McFadden test: Removes an alternative (supposed to be irrelevant) and check if β changes.

Testing the main assumption:

The Hausmann-McFadden test: Removes an alternative (supposed to be irrelevant) and check if β changes.

- Restricted model (a choice is removed) vs. unrestricted model (original)

Testing the main assumption:

The Hausmann-McFadden test: Removes an alternative (supposed to be irrelevant) and check if β changes.

- Restricted model (a choice is removed) vs. unrestricted model (original)
- if IIA holds, then unrestricted model has smaller variance.

$\Rightarrow$ χ^2*-test with smaller value indicatee IIA holds.*

Prediction testing

- **Predict outcome**

Prediction testing

- **Predict outcome**
  - predicted outcome/choice is the one with the highest probability/utility

Prediction testing

- **Predict outcome**
  - predicted outcome/choice is the one with the highest probability/utility
  - confusion matrix (Proportion of correct predictions: $\frac{\textit{sum of diagonal}}{\textit{N observations}}$)

Prediction testing

- **Predict outcome**
  - predicted outcome/choice is the one with the highest probability/utility
  - confusion matrix (Proportion of correct predictions: $\frac{\textit{sum of diagonal}}{\textit{N observations}}$)
- **Probability of all outcomes separately**: ROC curve and separation plots

Prediction testing

- **Predict outcome**
  - predicted outcome/choice is the one with the highest probability/utility
  - confusion matrix (Proportion of correct predictions: $\frac{\textit{sum of diagonal}}{\textit{N observations}}$)
- **Probability of all outcomes separately**: ROC curve and separation plots

$\Rightarrow$ *as in binomial regression, where you have one category vs. the rest*

Interpretation

All the possibilities of the binomial logit are open:

- The regression table

Interpretation

All the possibilities of the binomial logit are open:

- The regression table
- Predicted probabilities (and comparisons/scenarios) for each category

Interpretation

All the possibilities of the binomial logit are open:

- The regression table
- Predicted probabilities (and comparisons/scenarios) for each category
  - as with binomial logit, one line per category

Interpretation

All the possibilities of the binomial logit are open:

- The regression table
- Predicted probabilities (and comparisons/scenarios) for each category
    - as with binomial logit, one line per category
    - *cumulative* predicted probabilitites $\rightarrow$ illustrates tradeoffs

Interpretation

All the possibilities of the binomial logit are open:

- The regression table
- Predicted probabilities (and comparisons/scenarios) for each category
  - as with binomial logit, one line per category
  - *cumulative* predicted probabilitites → illustrates tradeoffs

⇒ *Remember reference cateogry is* $1-$ *the sum of all other probabilities*

Specific visual interpretations

If you have three categories (if $M = 3$)

- The three dimensional simplex

Specific visual interpretations

If you have three categories (if $M = 3$)

- The three dimensional simplex
- The ternary plot: a sort of scatterplot for predicted probabilities

Specific visual interpretations

If you have three categories (if $M = 3$)

- The three dimensional simplex
- The ternary plot: a sort of scatterplot for predicted probabilities

$\Rightarrow$ *Illustrates tradeoffs*

The conditional logit

From the chooser's perspectives

The conditional logit holds the chooser constant, and considers alternative choices

From the chooser's perspectives

The conditional logit holds the chooser constant, and considers alternative choices

- x refers to characteristics of the choice (not chooser)

From the chooser's perspectives

The conditional logit holds the chooser constant, and considers alternative choices

- x refers to characteristics of the choice (not chooser)
- Parallel regressions approach: a logit in a choice set

From the chooser's perspectives

The conditional logit holds the chooser constant, and considers alternative choices

- x refers to characteristics of the choice (not chooser)
- Parallel regressions approach: a logit in a choice set
- One set of parameters, no intercept

From the chooser's perspectives

The conditional logit holds the chooser constant, and considers alternative choices

- x refers to characteristics of the choice (not chooser)
- Parallel regressions approach: a logit in a choice set
- One set of parameters, no intercept
- Long data format (observation = choice in individual)

From the chooser's perspectives

The conditional logit holds the chooser constant, and considers alternative choices

- x refers to characteristics of the choice (not chooser)
- Parallel regressions approach: a logit in a choice set
- One set of parameters, no intercept
- Long data format (observation = choice in individual)

Mixing choosers and choices

The mixed conditional logit makes an interaction effect between choice-set variables and choice variables.

- Think hierarchical models with cross-level interactions