PLSC 504 – Fall 2017 Models for Nominal Outcomes

September 7, 2017

Nominal / Unordered Data

- Voter choice
- Voting in legislatures
- Joining international organizations
- Occupational choice, marketing, etc.
- Others?

Motivation: Discrete *Outcomes*

$$\Pr(Y_i = j) = P_{ij}$$

$$\sum_{j=1}^{J} P_{ij} = 1$$

$$P_{ij} = \exp(\mathbf{X}_i \boldsymbol{\beta}_j)$$

Motivation, continued

Rescale:

$$\Pr(Y_i = j) \equiv P_{ij} = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_j)}$$

Ensures

- $\Pr(Y_i = j) \in (0, 1)$
- $\sum_{j=1}^{J} \Pr(Y_i = j) = 1.0$

Identification

Constrain $\boldsymbol{\beta}_1 = \mathbf{0}$; then:

$$\Pr(Y_i = 1) = \frac{1}{1 + \sum_{j=2}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}'_j)}$$

$$\Pr(Y_i = j) = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}'_j)}{1 + \sum_{j=2}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}'_j)}$$

where $\boldsymbol{\beta}'_j = \boldsymbol{\beta}_j - \boldsymbol{\beta}_1$.

Alternative Motivation: Discrete *Choice*

$$U_{ij} = \mu_i + \epsilon_{ij}$$

$$\mu_i = \mathbf{X}_i \boldsymbol{\beta}_j$$

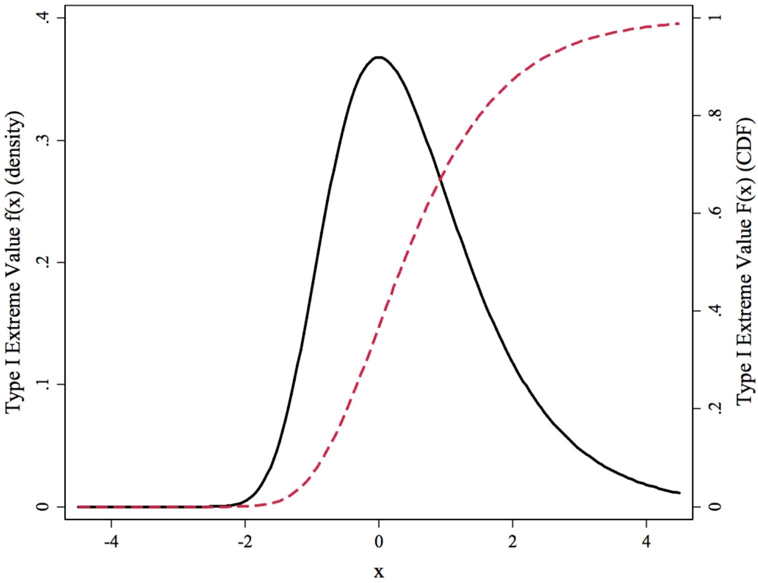
$$\begin{aligned}
\Pr(Y_i = j) &= \Pr(U_{ij} > U_{i\ell} \,\forall\, \ell \neq j \in J) \\
&= \Pr(\mu_i + \epsilon_{ij} > \mu_i + \epsilon_{i\ell} \,\forall\, \ell \neq j \in J) \\
&= \Pr(\mathbf{X}_i \boldsymbol{\beta}_j + \epsilon_{ij} > \mathbf{X}_i \boldsymbol{\beta}_\ell + \epsilon_{i\ell} \,\forall\, \ell \neq j \in J) \\
&= \Pr(\epsilon_{ij} - \epsilon_{i\ell} > \mathbf{X}_i \boldsymbol{\beta}_\ell - \mathbf{X}_i \boldsymbol{\beta}_j \,\forall\, \ell \neq j \in J)
\end{aligned}$$

Discrete Choice (continued)

$\epsilon \sim$???

- *Type I Extreme Value*
- Density: $f(\epsilon) = \exp[-\epsilon - \exp(-\epsilon)]$
- CDF: $\int f(\epsilon) \equiv F(\epsilon) = \exp[-\exp(-\epsilon)]$

Type I Extreme Value



$\rightarrow$ Model

$$
\begin{aligned}
\Pr(Y_i = j) &= \Pr(U_j > U_1, U_j > U_2, ...U_j > U_J) \\
&= \int f(\epsilon_j) \left[\int_{-\infty}^{\epsilon_{ij} + \mathbf{X}_i \boldsymbol{\beta}_j - \mathbf{X}_i \boldsymbol{\beta}_1} f(\epsilon_1) d\epsilon_1 \times \int_{-\infty}^{\epsilon_{ij} + \mathbf{X}_i \boldsymbol{\beta}_j - \mathbf{X}_i \boldsymbol{\beta}_2} f(\epsilon_2) d\epsilon_2 \times ... \right] d\epsilon_j \\
&= \int f(\epsilon_j) \times \exp[-\exp(\epsilon_{ij} + \mathbf{X}_i \boldsymbol{\beta}_j - \mathbf{X}_i \boldsymbol{\beta}_1)] \times \\
&\quad \exp[-\exp(\epsilon_{ij} + \mathbf{X}_i \boldsymbol{\beta}_j - \mathbf{X}_i \boldsymbol{\beta}_2)] \times ... d\epsilon_j \\
&= \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_j)}
\end{aligned}
$$

Define:

$$\begin{aligned}\delta_{ij} &= 1 \text{ if } Y_i = j, \\ &= 0 \text{ otherwise.}\end{aligned}$$

Then:

$$\begin{aligned}L_i &= \prod_{j=1}^{J} [\Pr(Y_i = j)]^{\delta_{ij}} \\ &= \prod_{j=1}^{J} \left[\frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_j)} \right]^{\delta_{ij}}\end{aligned}$$

So:
$$L = \prod_{i=1}^{N} \prod_{j=1}^{J} \left[\frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_j)} \right]^{\delta_{ij}}$$

and (of course):
$$\ln L = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \ln \left[\frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_j)} \right]$$

A (Descriptive) Example: 1992 Election

- 1992 National Election Study
- $Y \in \{\text{Bush} = 1, \text{Clinton} = 2, \text{Perot} = 3\}$
- $N = 1473$.
- $X =$ Party ID: $\{$"Strong Democrats" $= 1 \rightarrow$ "Strong Republicans" $= 7\}$

MNL: 1992 Election ("Baseline" = Perot)

```
> nes92.mlogit<-vglm(presvote~partyid, multinomial, nes92)
> summary(nes92.mlogit)

Call:
vglm(formula = presvote ~ partyid, family = multinomial, data = nes92)

Coefficients:
              Estimate Std. Error z value             Pr(>|z|)
(Intercept):1  -1.8152     0.2456   -7.39     0.00000000000014 ***
(Intercept):2   3.0273     0.1783   16.98 < 0.0000000000000002 ***
partyid:1       0.4827     0.0476   10.15 < 0.0000000000000002 ***
partyid:2      -0.6805     0.0478  -14.25 < 0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Number of linear predictors:  2
Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
Dispersion Parameter for multinomial family:   1
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of iterations: 5
```

MNL: 1992 Election ("Baseline" $=$ Bush)

```
> Bush.nes92.mlogit<-vglm(formula = presvote~partyid,
          family=multinomial(refLevel=1),data=nes92)
> summary(Bush.nes92.mlogit)

Coefficients:
              Estimate Std. Error z value              Pr(>|z|)
(Intercept):1   4.8425     0.2373   20.41 < 0.0000000000000002 ***
(Intercept):2   1.8152     0.2456    7.39     0.00000000000014 ***
partyid:1      -1.1632     0.0546  -21.32 < 0.0000000000000002 ***
partyid:2      -0.4827     0.0476  -10.15 < 0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Number of linear predictors:  2
Names of linear predictors: log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1])
Dispersion Parameter for multinomial family:   1
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of iterations: 5
```

MNL: 1992 Election (“Baseline” = Clinton)

```
> Clinton.nes92.mlogit<-vglm(formula=presvote~partyid,
                 family=multinomial(refLevel=2),data=nes92)
> summary(Clinton.nes92.mlogit)

Coefficients:
              Estimate Std. Error z value            Pr(>|z|)
(Intercept):1  -4.8425     0.2373   -20.4 <0.0000000000000002 ***
(Intercept):2  -3.0273     0.1783   -17.0 <0.0000000000000002 ***
partyid:1       1.1632     0.0546    21.3 <0.0000000000000002 ***
partyid:2       0.6805     0.0478    14.2 <0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Number of linear predictors:  2
Names of linear predictors: log(mu[,1]/mu[,2]), log(mu[,3]/mu[,2])
Dispersion Parameter for multinomial family:   1
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of iterations: 5
```

Coefficient Estimates and "Baselines"

		"Baseline" category		
		Clinton	Perot	Bush
Comparison	Clinton	–	-0.68	-1.16
Category	Perot	0.68	–	-0.48
	Bush	1.16	0.48	–

It is exactly the same as the multinomial logit model. Period.

Choice-Specific Covariates

```
> library(mlogit)
> colnames(nes92)<-c("caseid","presvote","partyid","FT.Bush",
  "FT.Clinton","FT.Perot")
> nes92$PVote<-factor(nes92$presvote,labels=c("Bush","Clinton","Perot"))
> nes92CL<-mlogit.data(nes92,shape="wide",choice="PVote",varying=4:6)
> head(nes92)

  caseid presvote partyid FT.Bush FT.Clinton FT.Perot      NA   PVote
1   3001        1       6      85         30        0    Bush    Bush
2   3002        1       7     100          0        0    Bush    Bush
3   3003        1       7      85         30       60    Bush    Bush
4   3005        2       6      40         60       60 Clinton Clinton
5   3006        2       2      30         70       50 Clinton Clinton
6   3007        2       1      15         70       50 Clinton Clinton
```

Conditional Logit

$$\Pr(Y_{ij} = 1) = \frac{\exp(\mathbf{Z}_{ij}\gamma)}{\sum_{j=1}^{J} \exp(\mathbf{Z}_{ij}\gamma)}$$

Combinations: $\mathbf{X}_i\boldsymbol{\beta}$ and $\mathbf{Z}_{ij}\gamma$

- "Fixed effects"
- Observation-specific **X**s
- Interactions...

CL in R : Estimation

```
> nes92.clogit<-mlogit(PVote~FT|partyid,data=nes92CL)
> summary(nes92.clogit)

Call:
mlogit(formula = PVote ~ FT | partyid, data = nes92CL, method = "nr",
    print.level = 0)

Frequencies of alternatives:
   Bush Clinton   Perot
  0.339   0.469   0.191

nr method
6 iterations, 0h:0m:0s
g'(-H)^-1g = 0.00293
successive function values within tolerance limits

Coefficients :
                    Estimate Std. Error t-value             Pr(>|t|)
Clinton:(intercept)  2.81272    0.26880   10.46 < 0.0000000000000002 ***
Perot:(intercept)    0.94353    0.28563    3.30              0.00096 ***
FT                   0.06299    0.00322   19.58 < 0.0000000000000002 ***
Clinton:partyid     -0.63187    0.06225  -10.15 < 0.0000000000000002 ***
Perot:partyid       -0.19212    0.05703   -3.37              0.00076 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log-Likelihood: -736
McFadden R^2:  0.519
Likelihood ratio test : chisq = 1590 (p.value = <0.0000000000000002)
```

Interpretation: Example Data Redux

- 1992 ANES ($N = 1473$)
- Variables:
  - `presvote`: 1=Bush, 2=Clinton, 3=Perot
  - `partyid`: (seven-point scale, 7=GOP)
  - `age` (in years)
  - `white` (naturally coded)
  - `female` (ditto)

Baseline MNL Results: 1992 Election

```
> NES.MNL<-vglm(presvote~partyid+age+white+female,data=BigNES92,
+           multinomial(refLevel=1))
> summaryvglm(NES.MNL)

Call:
vglm(formula = presvote ~ partyid + age + white + female, family = multinomial(refLevel = 1),
    data = BigNES92)

Coefficients:
              Estimate Std. Error z value              Pr(>|z|)
(Intercept):1  5.80665    0.44301   13.11 < 0.0000000000000002 ***
(Intercept):2  1.98008    0.52454    3.77              0.00016 ***
partyid:1     -1.13561    0.05486  -20.70 < 0.0000000000000002 ***
partyid:2     -0.50132    0.04870  -10.29 < 0.0000000000000002 ***
age:1         -0.00260    0.00514   -0.51              0.61276
age:2         -0.01556    0.00504   -3.09              0.00203 **
whiteWhite:1  -0.98908    0.31346   -3.16              0.00160 **
whiteWhite:2   0.87918    0.43605    2.02              0.04377 *
female:1      -0.12500    0.16895   -0.74              0.45936
female:2      -0.50928    0.16266   -3.13              0.00174 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Number of linear predictors:  2
Names of linear predictors: log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1])
Dispersion Parameter for multinomial family:   1
Residual deviance: 2107 on 2936 degrees of freedom
Log-likelihood: -1054 on 2936 degrees of freedom
Number of iterations: 5
```

MNL/CL: Model Fit

Global $\ln LR$ statistic Q tests:

$$\hat{\boldsymbol{\beta}} = \mathbf{0} \, \forall \, j, k$$

$$Q \sim \chi^2_{(J-1)(k-1)}$$

Test H: No Effect of age

```
> library(aod)
> wald.test(b=c(t(coef(NES.MNL))),Sigma=vcov(NES.MNL),Terms=c(5,6))

Wald test:
----------

Chi-squared test:
X2 = 11.0, df = 2, P(> X2) = 0.0042
```

Test H: No Difference – Clinton vs. Bush

```
> wald.test(b=c(t(coef(NES.MNL))),Sigma=vcov(NES.MNL),Terms=c(1,3,5,7,9))

Wald test:
----------

Chi-squared test:
X2 = 444.6, df = 5, P(> X2) = 0.0
```

Predicted Outcomes

```
> PickBush<-ifelse(fitted.values(NES.MNL)[,1]>fitted.values(NES.MNL)[,2]
         & fitted.values(NES.MNL)[,1]>fitted.values(NES.MNL)[,3], 1,0)
> PickWJC<-ifelse(fitted.values(NES.MNL)[,2]>fitted.values(NES.MNL)[,1]
         & fitted.values(NES.MNL)[,2]>fitted.values(NES.MNL)[,3], 2, 0)
> PickHRP<-ifelse(fitted.values(NES.MNL)[,3]>fitted.values(NES.MNL)[,1]
         & fitted.values(NES.MNL)[,3]>fitted.values(NES.MNL)[,2], 3, 0)

> OutHat<-PickBush+PickWJC+PickHRP
> table(BigNES92$presvote,OutHat)

   OutHat
      1   2   3
  1 415  77   8
  2  56 619  16
  3 135 133  14
```

- "Null" Model: $\left(\frac{691}{1473}\right) = 46.9\%$ correct.
- Estimated model: $\frac{(415+619+14)}{1473} = \frac{1048}{1473} =$ 71.2% correct.
- PRE $= \frac{1048-691}{1473-691} = \frac{357}{782} = 45.7\%$.
- Correct predictions: 90% Clinton, 83% Bush, 5% Perot.

Marginal Effects

$$\frac{\partial \Pr(Y_i = j)}{\partial X_k} = \Pr(Y_i = j|\mathbf{X})\left[\hat{\boldsymbol{\beta}}_{jk} - \sum_{j=1}^{J} \hat{\boldsymbol{\beta}}_{jk} \times \Pr(Y_i = j|\mathbf{X})\right]$$

Depends on:

- $\Pr\widehat{(Y_i = } j)$
- $\hat{\beta}_{jk}$
- $\sum_{j=1}^{J} \hat{\boldsymbol{\beta}}_{jk}$

See the end for (Stata) examples...

Odds ("Relative Risk") Ratios

$$\ln\left[\frac{\Pr(Y_i = j|\mathbf{X})}{\Pr(Y_i = j'|\mathbf{X})}\right] = \mathbf{X}(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{j'})$$

Setting $\hat{\boldsymbol{\beta}}_{j'} = \mathbf{0}$:

$$\ln\left[\frac{\Pr(Y_i = j|\mathbf{X})}{\Pr(Y_i = j'|\mathbf{X})}\right] = \mathbf{X}\hat{\boldsymbol{\beta}}_j$$

One-Unit Change in X_k:

$$RRR_{jk} = \exp(\boldsymbol{\beta}_{jk})$$

δ-Unit Change in X_k:

$$RRR_{jk} = \exp(\boldsymbol{\beta}_{jk} \times \delta)$$

Odds ("Relative Risk") Ratios

```
> mnl.or <- function(model) {
    coeffs <- c(t(coef(model)))
    lci <- exp(coeffs - 1.96 * diag(vcov(NES.MNL))^0.5)
    or <- exp(coeffs)
    uci <- exp(coeffs + 1.96* diag(vcov(NES.MNL))^0.5)
    lreg.or <- cbind(lci, or, uci)
    lreg.or
  }

> mnl.or(NES.MNL)
                    lci       or      uci
(Intercept):1 139.5398 332.5036 792.3088
(Intercept):2   2.5909   7.2433  20.2504
partyid:1       0.2885   0.3212   0.3577
partyid:2       0.5506   0.6057   0.6664
age:1           0.9874   0.9974   1.0075
age:2           0.9749   0.9846   0.9943
whiteWhite:1    0.2012   0.3719   0.6875
whiteWhite:2    1.0248   2.4089   5.6623
female:1        0.6337   0.8825   1.2289
female:2        0.4369   0.6009   0.8266
```

Odds Ratios: Interpretation

- A one unit increase in `partyid` corresponds to:
  - A decrease in the odds of a Clinton vote, versus a vote for Bush, of $\exp(-1.136) = 0.321$ (or about 68 percent), and
  - A decrease in the odds of a Perot vote, versus a vote for Bush, of $\exp(-0.501) = 0.606$ (or about 40 percent).
  - These are *large* decreases in the odds – not surprisingly, more Republican voters are *much* more likely to vote for Bush than for Perot or Clinton.
- Similarly, `female` voters are:
  - No more or less likely to vote for Clinton vs. Bush (OR=0.88), but
  - Roughly 40 percent less likely to have voted for Perot (OR=0.60).

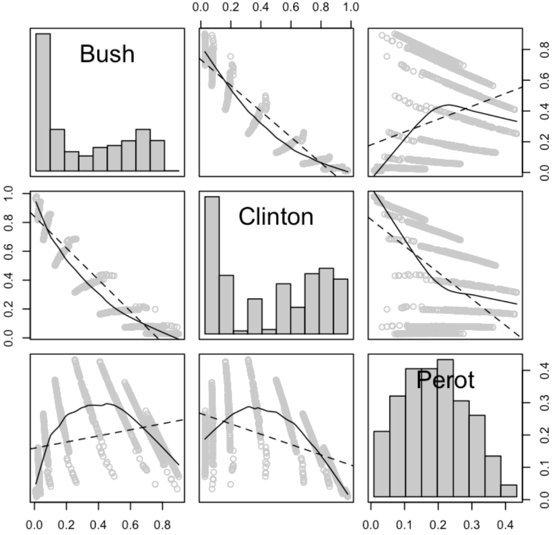
Predicted Probabilities

$$
\begin{aligned}
\Pr(\widehat{\texttt{presvote}_i} = \text{Bush}) &= \frac{\exp(\mathbf{X}_i \hat{\boldsymbol{\beta}}_{\text{Bush}})}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \hat{\boldsymbol{\beta}}_j)} \\
&= \frac{1}{1 + \sum_{j=2}^{J} \exp(\mathbf{X}_i \hat{\boldsymbol{\beta}}_j)}
\end{aligned}
$$

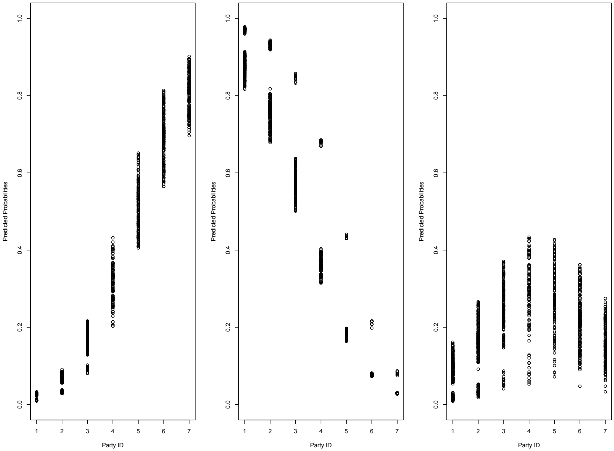
In-Sample Predicted Probabilities

```
> hats<-as.data.frame(fitted.values(NES.MNL))
> names(hats)[3]<-"Perot" # nice names...
> names(hats)[2]<-"Clinton"
> names(hats)[1]<-"Bush"
> attach(hats)

> library(car)
> scatterplot.matrix(~Bush+Clinton+Perot,
  diagonal="histogram",col=c("black","grey"))
```

In-Sample $\widehat{\text{Prs}}$



In-Sample $\widehat{\text{Pr}}$s vs. `partyid`

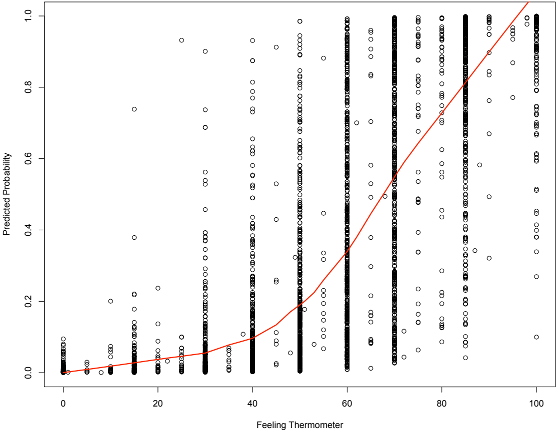


Conditional Logit: Example

```
> nes92.clogit<-mlogit(PVote~FT|partyid,data=nes92CL)
> summary(nes92.clogit)

Call:
mlogit(formula = PVote ~ FT | partyid, data = nes92CL, method = "nr",
    print.level = 0)

nr method
6 iterations, 0h:0m:0s
g'(-H)^-1g = 0.00293
successive function values within tolerance limits

Coefficients :
                    Estimate Std. Error t-value
Clinton:(intercept)  2.81272    0.26880   10.46
Perot:(intercept)    0.94353    0.28563    3.30
FT                   0.06299    0.00322   19.58
Clinton:partyid     -0.63187    0.06225  -10.15
Perot:partyid       -0.19212    0.05703   -3.37
                                Pr(>|t|)
Clinton:(intercept) < 0.0000000000000002 ***
Perot:(intercept)                0.00096 ***
FT                  < 0.0000000000000002 ***
Clinton:partyid     < 0.0000000000000002 ***
Perot:partyid                    0.00076 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log-Likelihood: -736
McFadden R^2:  0.519
Likelihood ratio test : chisq = 1590 (p.value = <0.0000000000000002)
```

Predicted Probabilities (In-Sample)

```
> CLhats<-predict(NES.CL,type="expected")
> plot(cldata$FT,CLhats,xlab="Feeling Thermometer",ylab="Predicted Probability")
> lines(lowess(CLhats~cldata$FT),lwd=2,col="red")
```



Independence of Irrelevant Alternatives

"An individual's choice does not depend on the availability or characteristics of unavailable alternatives."

IIA, Statistically

$$
\begin{aligned}
\frac{\Pr(Y_i = k)}{\Pr(Y_i = \ell)} &= \frac{\frac{\exp(\mathbf{X}_i\beta_k)}{\sum_{j=1}^{J}\exp(\mathbf{X}_i\beta_j)}}{\frac{\exp(\mathbf{X}_i\beta_\ell)}{\sum_{j=1}^{J}\exp(\mathbf{X}_i\beta_j)}} \\
&= \frac{\exp(\mathbf{X}_i\beta_k)}{\exp(\mathbf{X}_i\beta_\ell)} \\
&= \exp[\mathbf{X}_i(\beta_k - \beta_\ell)]
\end{aligned}
$$

Alternatively:

$$
\frac{\Pr(Y_i = k|S_J)}{\Pr(Y_i = \ell|S_J)} = \frac{\Pr(Y_i = k|S_M)}{\Pr(Y_i = \ell|S_M)} \; \forall \; k, \ell, J, M
$$

IIA, Intuitively

- Initially: $\Pr(\text{Car}) = \Pr(\text{Red Bus}) = 0.5$, $\frac{\Pr(\text{Car})}{\Pr(\text{Red Bus})} = 1$.
- Enter the Blue Bus...
  - Intuitively: $\Pr(\text{Car}) = 0.5$, $\Pr(\text{Red Bus}) = 0.25$, $\Pr(\text{Blue Bus}) = 0.25$
  - IIA: $\Pr(\text{Car}) = \Pr(\text{Red Bus}) = \Pr(\text{Blue Bus}) = 0.33$, or
  - $\Pr(\text{Car}) = \Pr(\text{Red Bus}) = 0.4$, $\Pr(\text{Blue Bus}) = 0.2$...
  - ...so long as $\frac{\Pr(\text{Car})}{\Pr(\text{Red Bus})} = 1$.

Why IIA?

$$\begin{aligned} U_{ij} &= \mu_{ij} + \epsilon_{ij} \\ &= \mathbf{X}_i \boldsymbol{\beta}_j + \epsilon_{ij} \end{aligned}$$

$$\begin{aligned} \Pr(Y_i = j) &= \Pr(U_{ij} > U_{i\ell}) \forall \ell \neq j \in J \\ &= \Pr(\mathbf{X}_i \boldsymbol{\beta}_j + \epsilon_{ij} > \mathbf{X}_i \boldsymbol{\beta}_\ell + \epsilon_{i\ell}) \forall \ell \neq j \in J \\ &= \Pr(\epsilon_{ij} - \epsilon_{i\ell} > \mathbf{X}_i \boldsymbol{\beta}_\ell - \mathbf{X}_i \boldsymbol{\beta}_j) \forall \ell \neq j \in J \end{aligned}$$

IIA Tests: Hausman/McFadden and Small/Hsiao

$$HM = (\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_u)'[\hat{\mathbf{V}}_r - \hat{\mathbf{V}}_u]^{-1}(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_u)$$

$$\widehat{HM} \sim \chi^2_{(J-2)k}$$

$$SH = -2\left[L_r(\hat{\boldsymbol{\beta}}^{AB}_u) - L_r(\hat{\boldsymbol{\beta}}^{B}_r)\right]$$

$$\widehat{SH} \sim \chi^2_{k_r}$$

IIA Freedom: Multinomial Probit

$\epsilon_{ij} \sim MVN(0, \Sigma)$, where:

$$\underset{J \times J}{\boldsymbol{\Sigma}} = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1J} \\ \vdots & \ddots & \vdots \\ \sigma_{J1} & \dots & \sigma_J^2 \end{bmatrix}$$

Define $\eta_{ij\ell} = \epsilon_{ij} - e_{i\ell}$. Then:

$$\begin{aligned} \Pr(Y_i = j) &= \Pr(\eta_{ij\ell} > \mathbf{X}_i\boldsymbol{\beta}_\ell - \mathbf{X}_i\boldsymbol{\beta}_j) \,\forall\, \ell \neq j \in J \\ &= \int_{-\infty}^{\mathbf{X}_i\boldsymbol{\beta}_1 - \mathbf{X}_i\boldsymbol{\beta}_j} \dots \int_{-\infty}^{\mathbf{X}_i\boldsymbol{\beta}_\ell - \mathbf{X}_i\boldsymbol{\beta}_j} \phi_J(\eta_{ij1}, \eta_{ij2}, \dots \eta_{ij\ell}) d\eta_{ij1}, \eta_{ij2}, \dots \eta_{ij\ell} \end{aligned}$$

MNP, Issues

- Identification: (Potentially) Fragile
- Estimation:
  - Always hard
  - Quadrature, or
  - Simulation (MCMC)

IIA Freedom: HEV

$$
\begin{aligned}
f(\epsilon_{ij}) &= \lambda(\epsilon_{ij}) \\
&= \frac{1}{\theta_j} \exp\left(-\frac{\epsilon_{ij}}{\theta_j}\right) \exp\left[-\exp\left(-\frac{\epsilon_{ij}}{\theta_j}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
F(\epsilon_{ij}) &= \Lambda(\epsilon_{ij}) \\
&= \int_{-\infty}^{z} f(\epsilon_{ij}) d\,\epsilon_{ij} \\
&= \exp\left[-\exp\left(-\frac{\epsilon_{ij}}{\theta_j}\right)\right]
\end{aligned}
$$

More HEV

$$\Pr(Y_i = j) = \int_{-\infty}^{\infty} \prod_{\ell \neq j} \Lambda\left(\frac{\mathbf{X}_i\boldsymbol{\beta}_j - \mathbf{X}_i\boldsymbol{\beta}_\ell + \epsilon_{ij}}{\theta_\ell}\right) \frac{1}{\theta_j} \lambda\left(\frac{\epsilon_{ij}}{\theta_j}\right) d\,\epsilon_{ij}$$

With $w = \frac{\epsilon_{ij}}{\theta_j}$:

$$\Pr(Y_i = j) = \int_{-\infty}^{\infty} \prod_{\ell \neq j} \Lambda\left(\frac{\mathbf{X}_i\boldsymbol{\beta}_j - \mathbf{X}_i\boldsymbol{\beta}_\ell + \theta_j w}{\theta_\ell}\right) \lambda(w) d\,w$$

MNL $\subset$ HEV

$\theta_j = 1 \ \forall \ j \rightarrow$

$$\Pr(Y_i = j) = \int_{-\infty}^{\infty} \prod_{\ell \neq j} \Lambda(\mathbf{X}_i \beta_j - \mathbf{X}_i \beta_\ell + \epsilon_{ij}) \lambda(\epsilon_{ij}) d\,\epsilon_{ij}$$

IIA Freedom: "Mixed Logit"

$$U_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \epsilon_{ij},$$

$$\epsilon_{ij} = \eta_i + \xi_{ij}$$

$$\Pr(Y_i = j|\eta) \equiv \Pr(Y_{ij} = 1|\eta) = \frac{\exp(\mathbf{X}_{ij}\boldsymbol{\beta} + \eta_i)}{\sum_{j=1}^{J} \exp(\mathbf{X}_{ij}\boldsymbol{\beta} + \eta_i)}$$

What to do with the ηs?

Assume:

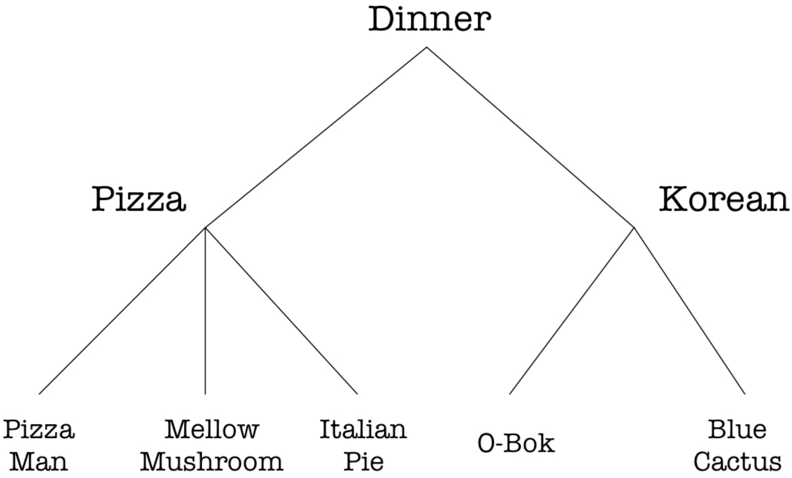
$$\eta_i \sim g(\mathbf{0}, \boldsymbol{\Omega})$$

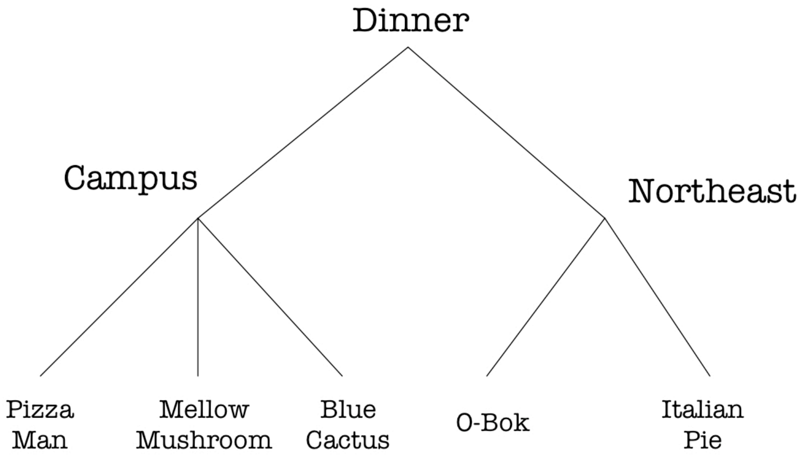
Yields:

$$\Pr(Y_i = j) = \int \left[\frac{\exp(\mathbf{X}_{ij}\boldsymbol{\beta} + \eta_i)}{\sum_{j=1}^{J} \exp(\mathbf{X}_{ij}\boldsymbol{\beta} + \eta_i)} \right] g(\eta|\boldsymbol{\Omega})\, d\eta$$

Nested Logit

- "Nested" choices
- A priori information about "subsets"
- IIA holds *within* (but not *across*) subsets...


Dinner
Pizza
Korean
Pizza Man
Mellow Mushroom
Italian Pie
O-Bok
Blue Cactus


Dinner
Campus
Northeast
Pizza Man
Mellow Mushroom
Blue Cactus
O-Bok
Italian Pie

Example: 2002 Swedish Election ($N = 6610$)

```
> summary(Sweden)

          partychoice        female            union           leftright    
 Conservatives   :1469   Min.   :0.0000   Min.   :1.000   Min.   :1.000  
 Liberals        :1212   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:2.000  
 Social Democrats:2975   Median :0.0000   Median :3.000   Median :3.000  
 Left Party      : 954   Mean   :0.4882   Mean   :2.709   Mean   :2.868  
                         3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
                         Max.   :1.0000   Max.   :4.000   Max.   :5.000  
      age       
 Min.   :17.00  
 1st Qu.:29.00  
 Median :42.00  
 Mean   :42.93  
 3rd Qu.:55.00  
 Max.   :90.00  
```

Swedish Election: MNL

```
> library(mlogit)
> Sweden.Long<-mlogit.data(Sweden,choice="partychoice",shape="wide")
> Sweden.MNL<-mlogit(partychoice~1|female+union+leftright+age,data=Sweden.Long)
> summary(Sweden.MNL)

Frequencies of alternatives:
   Conservatives       Left Party         Liberals Social Democrats
         0.22224          0.14433          0.18336          0.45008

Coefficients :
                               Estimate Std. Error  t-value  Pr(>|t|)
altLeft Party                13.3907039  0.3788540  35.3453 < 2.2e-16 ***
altLiberals                   4.4121638  0.2928137  15.0682 < 2.2e-16 ***
altSocial Democrats          11.3821332  0.3289066  34.6060 < 2.2e-16 ***
altLeft Party:female          0.7211951  0.1218437   5.9190 3.239e-09 ***
altLiberals:female            0.5585172  0.0848597   6.5817 4.652e-11 ***
altSocial Democrats:female    0.3881456  0.0945266   4.1062 4.022e-05 ***
altLeft Party:union          -0.4334637  0.0513499  -8.4414 < 2.2e-16 ***
altLiberals:union            -0.0563136  0.0388720  -1.4487 0.1474228
altSocial Democrats:union    -0.4145682  0.0408153 -10.1572 < 2.2e-16 ***
altLeft Party:leftright      -4.0917135  0.0930610 -43.9681 < 2.2e-16 ***
altLiberals:leftright        -1.1274488  0.0593125 -19.0086 < 2.2e-16 ***
altSocial Democrats:leftright -2.7555009  0.0719411 -38.3022 < 2.2e-16 ***
altLeft Party:age            -0.0277444  0.0038808  -7.1491 8.737e-13 ***
altLiberals:age              -0.0064185  0.0025768  -2.4909 0.0127410 *
altSocial Democrats:age      -0.0105052  0.0029196  -3.5982 0.0003204 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log-Likelihood: -5627.5
McFadden R^2:  0.33693
Likelihood ratio test : chisq = 5719 (p.value=< 2.22e-16)
```

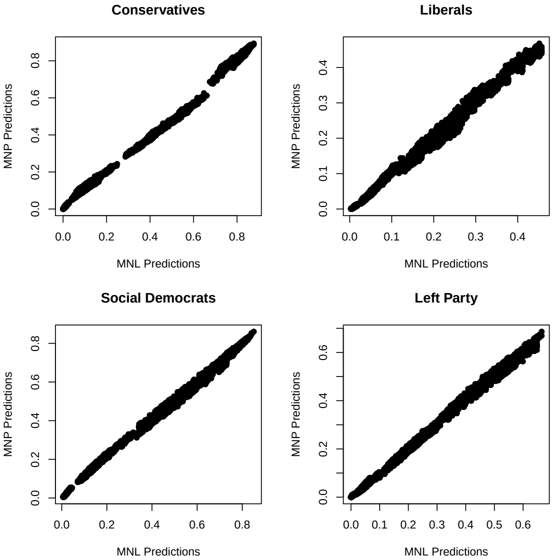
Hausman-McFadden IIA Test

```
> # Restricted model (omitting Social Democrats)
> Sweden.MNL.Restr<-mlogit(partychoice~1|female+union+leftright+age,
+ Sweden.Long,alt.subset=c("Conservatives","Liberals","Left Party"))
>
> hmftest(Sweden.MNL,Sweden.MNL.Restr)

Hausman-McFadden test

data:  Sweden.Long
chisq = 19.1137, df = 10, p-value = 0.03884
alternative hypothesis: IIA is rejected
```

Swedish Election: MNP

```
> library(MNP)
> Sweden.MNP<-mnp(partychoice~female+union+leftright+age, data=Sweden)
> summary(Sweden.MNP)

Coefficients:
                                  mean  std.dev.       2.5%      97.5%
(Intercept):Liberals          3.964677  0.879442  0.983572      4.669
(Intercept):Social Democrats  7.993453  1.495732  3.986961      9.812
(Intercept):Left Party       10.342468  2.082971  4.845935     12.714
female:Liberals               0.293136  0.046373  0.204654      0.382
female:Social Democrats       0.290311  0.079166  0.124746      0.447
female:Left Party             0.613163  0.163673  0.289974      0.944
union:Liberals               -0.083366  0.036782 -0.140052      0.024
union:Social Democrats       -0.275696  0.059260 -0.369943     -0.145
union:Left Party             -0.346922  0.087131 -0.489992     -0.148
leftright:Liberals           -0.913247  0.168331 -1.045781     -0.350
leftright:Social Democrats   -1.920076  0.362403 -2.371245     -0.977
leftright:Left Party         -3.409277  0.750701 -4.308455     -1.576
age:Liberals                 -0.003350  0.001490 -0.006264 -0.000409
age:Social Democrats         -0.007171  0.002630 -0.012327     -0.002
age:Left Party               -0.025595  0.007323 -0.039641     -0.011

Covariances:
                                    mean std.dev.   2.5%  97.5%
Liberals:Liberals                 1.0000   0.0000 1.0000  1.000
Liberals:Social Democrats         1.4083   0.3925 0.2116  1.830
Liberals:Left Party               2.4450   1.0779 0.6731  3.988
Social Democrats:Social Democrats 2.6696   0.9215 0.5630  3.898
Social Democrats:Left Party       4.4852   2.1846 0.3521  7.524
Left Party:Left Party             9.4811   5.0787 1.1682 17.095

Base category: Conservatives
Number of alternatives: 4
Number of observations: 6610
Number of estimated parameters: 20
Number of stored MCMC draws: 5000
```

Or, How I Stopped Worrying and Learned To Love MNL...



Software

Model	Stata	SAS	R
Multinomial Logit	`mlogit`	`proc catmod`	`vglm, mlogit, multinom`*
Conditional Logit	`clogit`	`proc mdc`	`clogit, mlogit`
Multinomial Probit	`mprobit` / `asmprobit`	`proc mdc`	`mnp`*
Heteroscedastic Extreme Value	No(?)	`proc mdc`	`mlogit`
Mixed Logit**	`mixlogit`	`proc mdc`	`mlogit`
Nested Logit	`nlogit`	`proc mdc`	`mlogit`

* See also `bayesm`, `MCMCpack`.

** Can also be estimated using GAUSS.