PLSC 503 – Spring 2022
Regression Models for Nominal and Ordinal Outcomes

April 13, 2022

Motivation: Discrete *Outcomes*

Outcome variable has $J > 2$ *unordered* categories:

$$Y_i \in \{1, 2, ...J\}$$

Write:

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

Means that:

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

And set:

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

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

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

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*

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

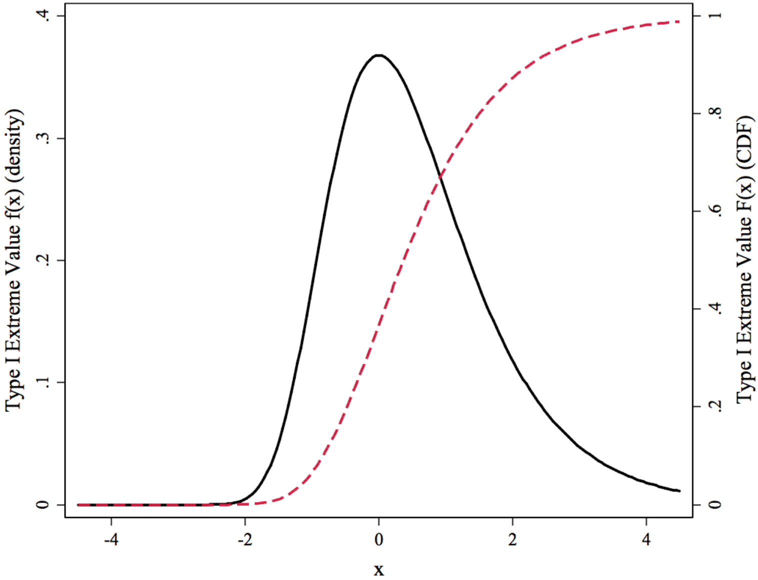
$$\mu_i = \mathbf{X}_i \boldsymbol{\beta}_j$$

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

$\epsilon \sim$???

- *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



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

Define:

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

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  1.4e-13 ***
(Intercept):2   3.0273     0.1783   16.98  < 2e-16 ***
partyid:1       0.4827     0.0476   10.15  < 2e-16 ***
partyid:2      -0.6805     0.0478  -14.25  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of Fisher scoring iterations: 5
No Hauck-Donner effect found in any of the estimates
Reference group is level  3  of the response
```

MNL: 1992 Election ("Baseline" = Bush)

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

Call:
vglm(formula = presvote ~ partyid, family = multinomial(refLevel = 1),
    data = nes92)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1   4.8425     0.2373   20.41  < 2e-16 ***
(Intercept):2   1.8152     0.2456    7.39  1.4e-13 ***
partyid:1      -1.1632     0.0546  -21.32  < 2e-16 ***
partyid:2      -0.4827     0.0476  -10.15  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Names of linear predictors: log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1])
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of Fisher scoring iterations: 5
No Hauck-Donner effect found in any of the estimates
Reference group is level  1  of the response
```

MNL: 1992 Election ("Baseline" = Clinton)

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

Call:
vglm(formula = presvote ~ partyid, family = multinomial(refLevel = 2),
    data = nes92)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1  -4.8425     0.2373   -20.4   <2e-16 ***
(Intercept):2  -3.0273     0.1783   -17.0   <2e-16 ***
partyid:1       1.1632     0.0546    21.3   <2e-16 ***
partyid:2       0.6805     0.0478    14.2   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Names of linear predictors: log(mu[,1]/mu[,2]), log(mu[,3]/mu[,2])
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of Fisher scoring iterations: 5
Reference group is level  2  of the response
```

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: Data Structure

```
> nes92CL<-mlogit.data(nes92,shape="wide",choice="PVote",varying=4:6)
> head(nes92CL,6)
~~~~~~~~
 first 6 observations out of 4419
~~~~~~~~
  caseid presvote partyid PVote     alt  FT chid    idx
1   3001        1       6  TRUE    Bush  85    1 1:Bush
2   3001        1       6 FALSE Clinton  30    1 1:nton
3   3001        1       6 FALSE   Perot   0    1 1:erot
4   3002        1       7  TRUE    Bush 100    2 2:Bush
5   3002        1       7 FALSE Clinton   0    2 2:nton
6   3002        1       7 FALSE   Perot   0    2 2:erot

~~~ indexes ~~~~
  chid     alt
1    1    Bush
2    1 Clinton
3    1   Perot
4    2    Bush
5    2 Clinton
6    2   Perot
indexes:  1, 2
```

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

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

- "Fixed effects" (choice-specific intercepts), plus
- Observation-specific $\mathbf{X}$s, plus
- 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` $\in$ {Bush, Clinton, Perot} (factor)
  - `presvote`: 1=Bush, 2=Clinton, 3=Perot (numeric)
  - `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  < 2e-16 ***
(Intercept):2  1.98008    0.52454    3.77  0.00016 ***
partyid:1     -1.13561    0.05486  -20.70  < 2e-16 ***
partyid:2     -0.50132    0.04870  -10.29  < 2e-16 ***
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

Names of linear predictors: log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1])
Residual deviance: 2107 on 2936 degrees of freedom
Log-likelihood: -1054 on 2936 degrees of freedom
Number of Fisher scoring iterations: 5
No Hauck-Donner effect found in any of the estimates
Reference group is level  1  of the response
```

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
```

In-Sample 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
```

$\rightarrow$ PRE

- "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.

Interpretation: 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:

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

Available for `-multinom-` (in the `-nnet-` package) via the `-margins-` package...

Marginal Effects: Illustrated

```
> Re-fit the model using -multinom-:
>
> BigNES92$PresVote<-cut(BigNES92$presvote,3,labels=c("Bush","Clinton","Perot")
> BigNES92$White<-ifelse(BigNES92$white=="White",1,0) # numeric
> MNL.alt<-multinom(PresVote~partyid+age+White+female,data=BigNES92,
+                   Hess=TRUE)
# weights:  18 (10 variable)
initial  value 1618.255901
iter  10 value 1077.315546
final  value 1053.650587
converged

> summary(marginal_effects(MNL.alt))
  dydx_partyid       dydx_age           dydx_White         dydx_female
 Min.   :0.0104   Min.   :0.00003   Min.   :-0.1482   Min.   :0.0013
 1st Qu.:0.0578   1st Qu.:0.00032   1st Qu.:-0.0608   1st Qu.:0.0125
 Median :0.1069   Median :0.00093   Median : 0.0190   Median :0.0344
 Mean   :0.1060   Mean   :0.00130   Mean   :-0.0044   Mean   :0.0450
 3rd Qu.:0.1490   3rd Qu.:0.00234   3rd Qu.: 0.0402   3rd Qu.:0.0801
 Max.   :0.2612   Max.   :0.00329   Max.   : 0.1805   Max.   :0.1093
```

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
```

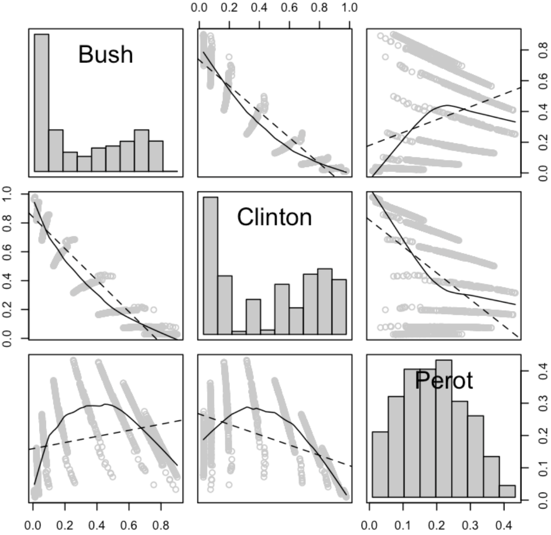
- A one unit increase in `partyid` corresponds to:
  - A decrease in the odds of a Clinton vote, versus a vote for Bush, of $\exp(-1.136) = 0.321$ (or about 68 percent), and
  - A decrease in the odds of a Perot vote, versus a vote for Bush, of $\exp(-0.501) = 0.606$ (or about 40 percent).
  - These 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).

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

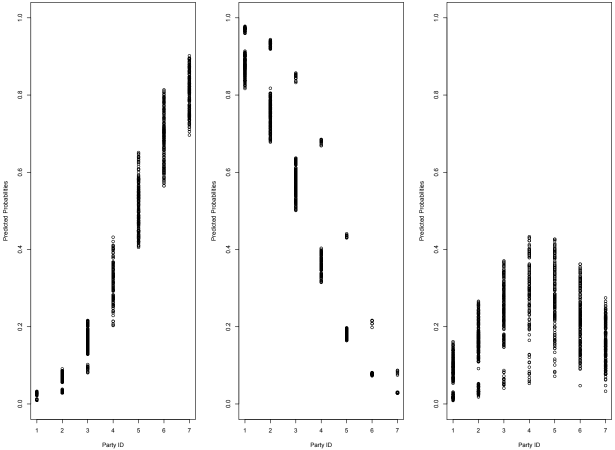
In-Sample Predicted Probabilities

```
> 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"))
```

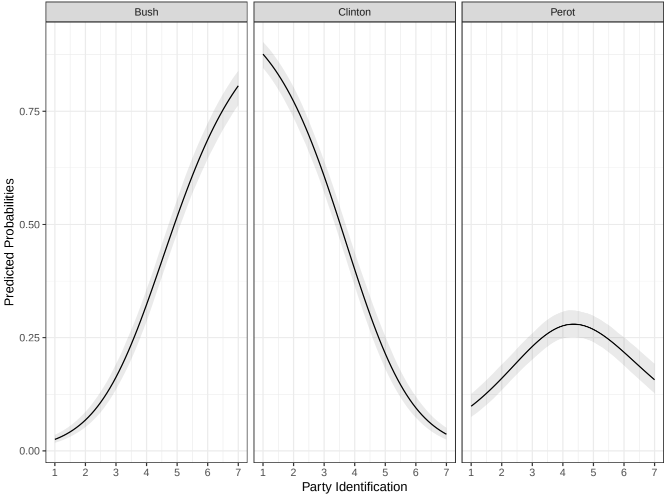
In-Sample $\widehat{\Pr}$s



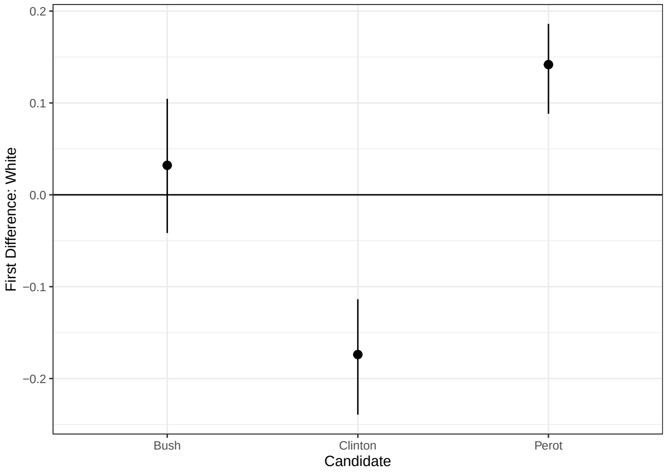
In-Sample $\widehat{\Pr}$s vs. `partyid`



Out-Of-Sample Predictions (using `MNLpred`)



OOS First Differences (using `MNLpred`)

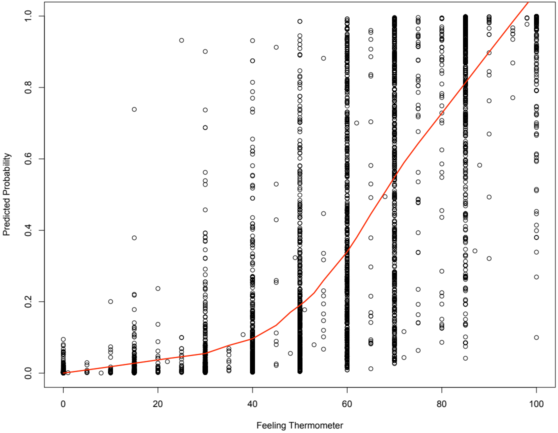


Conditional Logit: Example

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

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

Frequencies of alternatives:choice
   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 z-value Pr(>|z|)
(Intercept):Clinton  2.81272    0.26880   10.46  < 2e-16 ***
(Intercept):Perot    0.94353    0.28563    3.30  0.00096 ***
FT                   0.06299    0.00322   19.58  < 2e-16 ***
partyid:Clinton     -0.63187    0.06225  -10.15  < 2e-16 ***
partyid:Perot       -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 = <2e-16)
```

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")
```



Other Topics (for PLSC 504)

- "Independence of Irrelevant Alternatives"
- $\rightarrow$ Multinomial Probit
- $\rightarrow$ Heteroscedastic Extreme Value model
- "Mixed" Logit
- Nested Logit