PLSC 504 – Fall 2024

Regression Models for Nominal-Level Responses

September 11, 2024

Nominal 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

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

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: $\int f(\epsilon) \equiv F(\epsilon) = \exp[-\exp(-\epsilon)]$
- $\rightarrow$ Multinomial Logit

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):

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

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

Conditional Logit (CL)

CL with choice-varying predictors $\mathbf{Z}_{ij}\gamma$ is:

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

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

- "Fixed effects" for each possible outcome / choice
- Observation-specific $\mathbf{X}$s
- Interactions...

MNL and CL: Practical Things

The PLSC 503 slides and code include some additional detail, plus a running example (the three-candidate 1992 U.S. presidential election), with discussions of:

- Model estimation (including choosing the baseline/reference outcome),
- Model interpretation and discussion (odds ratios, predicted probabilities, etc.),
- Model fit, and
- Diagnostics.

I've included most of the code for those examples in today's code as well.

Independence of Irrelevant Alternatives ("IIA")

"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 requires that $\frac{\Pr(\text{Car})}{\Pr(\text{Red Bus})} = 1$.
    - So, that could be $\Pr(\text{Car}) = \Pr(\text{Red Bus}) = \Pr(\text{Blue Bus}) = 0.33$, or
    - $\Pr(\text{Car}) = \Pr(\text{Red Bus}) = 0.4$ and $\Pr(\text{Blue Bus}) = 0.2$...

Why IIA?

Random utility model:

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

... means that:

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

IIA Tests: Hausman/McFadden and Small/Hsiao

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

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

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

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

IIA Freedom: Multinomial Probit

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

$$\underset{J \times J}{\boldsymbol{\Sigma}} = \left[\begin{array}{ccc} \sigma_1^2 & \ldots & \sigma_{1J} \\ \vdots & \ddots & \vdots \\ \sigma_{J1} & \ldots & \sigma_J^2 \end{array} \right]$$

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

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

MNP: Issues and Estimation

- Identification: (Potentially) Fragile
- Estimation:
  - Always hard
  - Via "GHK" algorithm, or
  - Gaussian quadrature, or
  - Simulation (MCMC) (preferred)
- Software:
  - `mlogit` with `probit = TRUE` (Geweke-Hajivassiliou-Keane algorithm)
  - `MNP` package (Bayesian/MCMC)
  - `endogMNP` package (Bayesian with endogenous switching)
  - Others?

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

Means:

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

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

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

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?

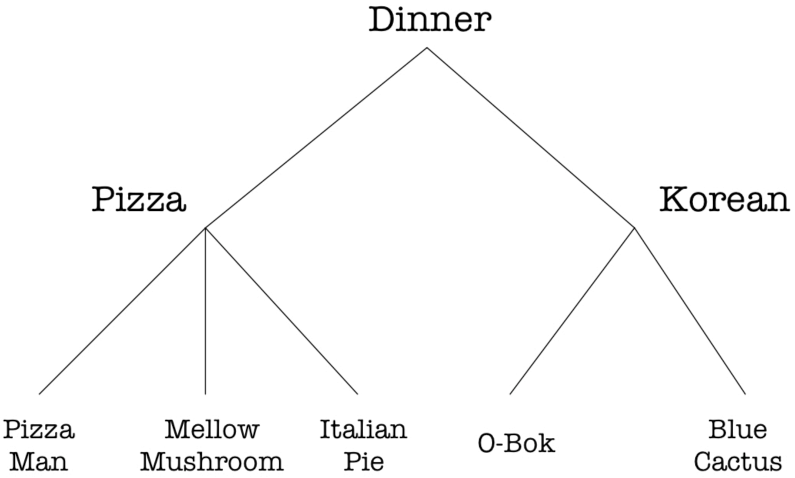
Assume:

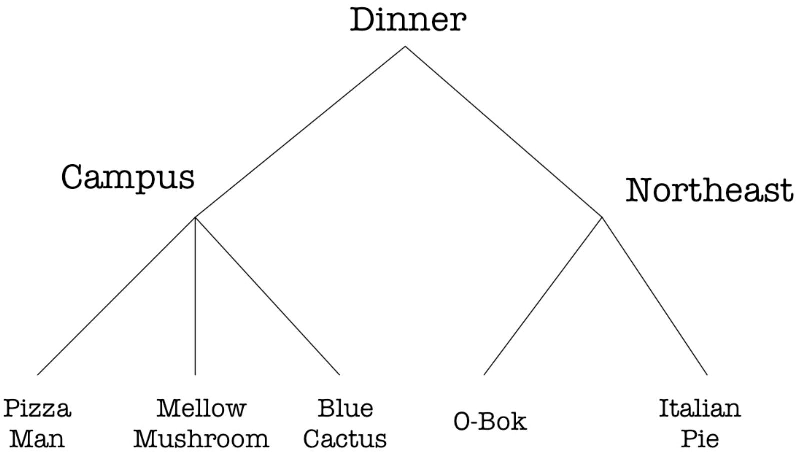
$$\eta_i \sim g(\mathbf{0}, \boldsymbol{\Omega})$$

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" 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
```

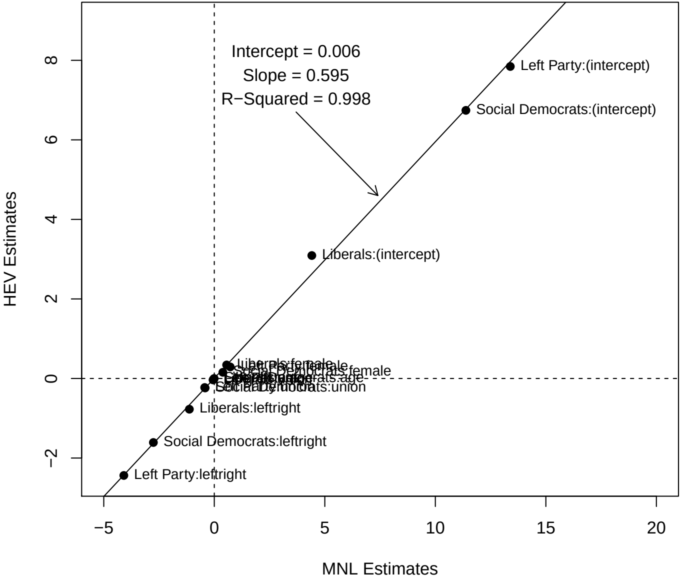
Swedish Election: HEV

```
> Sweden.Het<-mlogit(partychoice~1|female+union+leftright+
+                    age,data=Sweden.Long,heterosc=TRUE)
> summary(Sweden.Het)

Coefficients :
                              Estimate Std. Error z-value Pr(>|z|)
Left Party:(intercept)         7.84569    0.42849   18.31  < 2e-16 ***
Liberals:(intercept)           3.09199    0.30607   10.10  < 2e-16 ***
Social Democrats:(intercept)   6.74242    0.32038   21.04  < 2e-16 ***
Left Party:female              0.29096    0.08057    3.61   0.0003 ***
Liberals:female                0.34113    0.06510    5.24  1.6e-07 ***
Social Democrats:female        0.15572    0.05718    2.72   0.0065 **
Left Party:union              -0.22645    0.03704   -6.11  9.7e-10 ***
Liberals:union                -0.03498    0.02685   -1.30   0.1926
Social Democrats:union        -0.23786    0.03319   -7.17  7.8e-13 ***
Left Party:leftright          -2.43814    0.17450  -13.97  < 2e-16 ***
Liberals:leftright            -0.77255    0.04629  -16.69  < 2e-16 ***
Social Democrats:leftright    -1.60927    0.09462  -17.01  < 2e-16 ***
Left Party:age                -0.01612    0.00338   -4.77  1.9e-06 ***
Liberals:age                  -0.00200    0.00176   -1.14   0.2543
Social Democrats:age          -0.00267    0.00175   -1.53   0.1258
sp.Left Party                  0.90017    0.14304    6.29  3.1e-10 ***
sp.Liberals                    0.59981    0.09925    6.04  1.5e-09 ***
sp.Social Democrats            0.69163    0.10197    6.78  1.2e-11 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log-Likelihood: -5840
McFadden R^2:  0.312
Likelihood ratio test : chisq = 5300 (p.value = <2e-16)
```

$\hat{\beta}$s: MNL vs. HEV



Tests:

```
> MNL.HEV.Wald <- waldtest(Sweden.Het, heterosc = FALSE) # Wald test
> MNL.HEV.Wald

 Wald test

data:  homoscedasticity
chisq = 20, df = 3, p-value = 0.0004

> MNL.HEV.LR <- lrtest(Sweden.Het)         # LR test
> MNL.HEV.LR
Likelihood ratio test

Model 1: partychoice ~ 1 | female + union + leftright + age
Model 2: partychoice ~ 1 | female + union + leftright + age
  #Df LogLik Df Chisq Pr(>Chisq)
1  18  -5836
2  15  -5627 -3   416     <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> MNL.HEV.Score <- scoretest(Sweden.MNL, heterosc = TRUE)   # score test
> MNL.HEV.Score

 score test

data:  heterosc = TRUE
chisq = 20, df = 3, p-value = 0.00002
alternative hypothesis: heteroscedastic model
```

Swedish Election: MNP

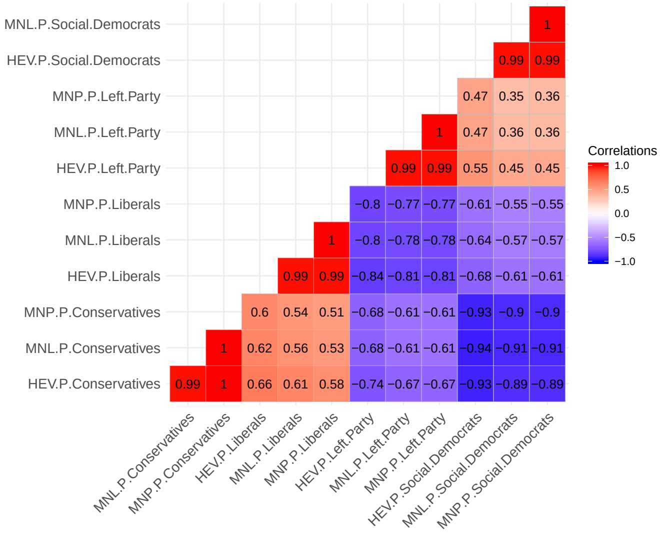
```
> library(MNP)
> Sweden.MNP<-mnp(partychoice~female+union+leftright+age, data=Sweden)
> summary(Sweden.MNP)

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
```

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`*, `mlogit`
Heteroscedastic Extreme Value	No(?)	`proc mdc`	`mlogit`
Mixed Logit	`mixlogit`	`proc mdc`	`mlogit`
Nested Logit	`nlogit`	`proc mdc`	`mlogit`

* See also `bayesm`.

Things To Read

- Bhat, Chandra R. 1995. "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice." *Transportation Research Part B: Methodological* 29(6):471-83.
- Colonescu, Constantin. 2016. *Principles of Econometrics with R*. Chapter 16: "Qualitative and LDV Models." Available here.
- Hensher, David A., and William H. Greene. 2002. "Specification and Estimation of the Nested Logit Model: Alternative Normalisations." *Transportation Research Part B* 36:1-17.
- Imai, Kosuke, and D. A. van Dyk. 2005. "A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation." *Journal of Econometrics* 124:311-334.
- McFadden, Daniel. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics* 3:303-28.
- Patty, John W., and Elizabeth M. Penn. 2019. "A Defense of Arrow's Independence of Irrelevant Alternatives." *Public Choice* 179:145-164.
- Sarrias, Mauricio, and Ricardo Daziano. 2017. "Multinomial Logit Models with Continuous and Discrete Individual Heterogeneity in R: The Gmnl Package." *Journal of Statistical Software, Articles* 79(2):1-46.
- Seshadri, Arjun, and Johan Ugander. 2020. "Fundamental Limits of Testing the Independence of Irrelevant Alternatives in Discrete Choice." Working paper, arXiv:2001.07042.
- Train, Kenneth. 2009. *Discrete Choice Methods with Simulation.* New York: Cambridge University Press.