Harvard Department of Government 2003
Faraway Chapter 7, Multinomial Data

JEFF GILL
Visiting Professor, Fall 2024

Bureaucratic Politics Example

- Contains *every* federal political appointee to full-time positions requiring Senate confirmation from November, 1964 through December, 1984 (collected by Mackenzie and Light, ICPSR Study Number 8458, Spring 1987).
- The survey queries various aspects of the Senate confirmation process, acclamation to running an agency or program, and relationships with other functions of government.
- The authors needed to preserve anonymity so they embargoed some variables and randomly sampled 1,500 down to 512.
- These latter issues are dealt with in Gill and Casella (JASA 2009).

Bureaucratic Politics Example

- **Outcome Variable:** `stress` as a surrogate measure for self-perceived effectiveness and job-satisfaction, measured as a five-point scale from "not stressful at all" to "very stressful."
- **Explanatory Variables:**
  - `Government Experience`,
  - `Ideology`,
  - `Committee Relationship`,
  - `Career.Exec-Compet`,
  - `Career.Exec-Liaison/Bur`,
  - `Career.Exec-Liaison/Cong`,
  - `Career.Exec-Day2day`,
  - `Career.Exec-Diff`,
  - `Confirmation Preparation`,
  - `Hours/Week`,
  - `President Orientation`.

Background

- The multinomial distribution is an extension of the binomial where the outcome is allowed to take on more than two values.
- Define Y_i as the *nominal* random variable taking on values $1, 2, \ldots, J$.
- Let $p_{ij} = p(Y_i = j)$ with the requirement that $\sum_{j=1}^{J} p_{ij} = 1$.
- Further define Y_{ij} as the number of observations falling into outcome j for case i.
- For *Grouped Data* types, these are cell counts where $n_i = \sum_j Y_{ij}$.
- For *Ungrouped Data* types, we have the restriction that $n_i = 1$ for exactly one outcome and $n_i = 0$ for the rest.
- The PMF is then given by:

$$p(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{ij}) = \frac{n_i}{y_{i1}! \cdots y_{iJ}!} p_{i1}^{y_{i1}} \cdots p_{iJ}^{y_{iJ}}$$

- The big distinction in this chapter: *ordered versus unordered data.*

Anderson's Typology of Ordinal Data

(JRSS-B, 1984, 1-30). Two scenarios:

1. *Grouped Continuous.*
   - Data originally measured on an interval or near-interval scale.
   - Later grouped for: convenience, compatability, or empirical reasons.
2. *Assessed Ordered.*
   - Categories exist in the original data collection effort.
   - Most common source: survey assessments.
3. Classic Reference:
   Zavoina, R., and W. McElvey. 1975.
   "A Statistical Model for the Analysis of Ordinal Level Dependent Variables."
   Journal of Mathematical Sociology (Summer), 103-20.

Threshold Approach for Ordinal Models

- $\exists$ $\mathbf{X}$, a matrix of explanatory variables.
- Y observed on ordered/recorded on ordered categories: $Y \in [1, \ldots, k]$.
- Y assumed to be produced by an unobserved (latent) variable U for assessed ordered case, or Y but inconvient U for grouped continuous case.
- U is continuous on $\Re$ for now (truncated later).
- The "response mechanism" for the r^{th} category: $Y = r \Longleftrightarrow \theta_{r-1} < U < \theta_r$
- This requires there to be thresholds on $\Re$ (no intercept):

$$\mathbf{U}_i: \;\; \theta_0 \underset{c=1}{\Longleftrightarrow} \theta_1 \underset{c=2}{\Longleftrightarrow} \theta_2 \underset{c=3}{\Longleftrightarrow} \theta_3 \ldots \theta_{C-1} \underset{c=C}{\Longleftrightarrow} \theta_C$$

- The vector of (unseen) utilities across individuals in the sample, $\mathbf{U}$, is determined by a linear additive specification of explanatory variables: $\mathbf{U} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}$, where $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_p]$ does not depend on the θ_j, and $\mathbf{E} \sim F_{\mathbf{E}}$.
- Some authors prefer a minus sign in front of $\mathbf{X}_i\boldsymbol{\beta}$, but the model defined here does not as is also the case with the `R` function `polr`: "$logit P(Y <= k|x) = zeta\_k - eta$".

Threshold Approach for Ordinal Models

- For the observed vector $\mathbf{Y}$:

$$\begin{aligned} p(\mathbf{Y} \le r|\mathbf{X}) &= p(\mathbf{U} \le \theta_r) = p(\mathbf{X}\boldsymbol{\beta} + \mathbf{E} \le \theta_r) \\ &= p(\mathbf{E} \le \theta_r - \mathbf{X}\boldsymbol{\beta}) = F_{\mathbf{E}}(\theta_r - \mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

- This is called the *cumulative model* because:

$$p(\mathbf{Y} \le \theta_r|\mathbf{X}) = p(\mathbf{Y} = 1|\mathbf{X}) + p(\mathbf{Y} = 2|\mathbf{X}) + \ldots + p(\mathbf{Y} = r|\mathbf{X})$$

- A logistic distributional assumption on the errors produces the ordered logit specification:

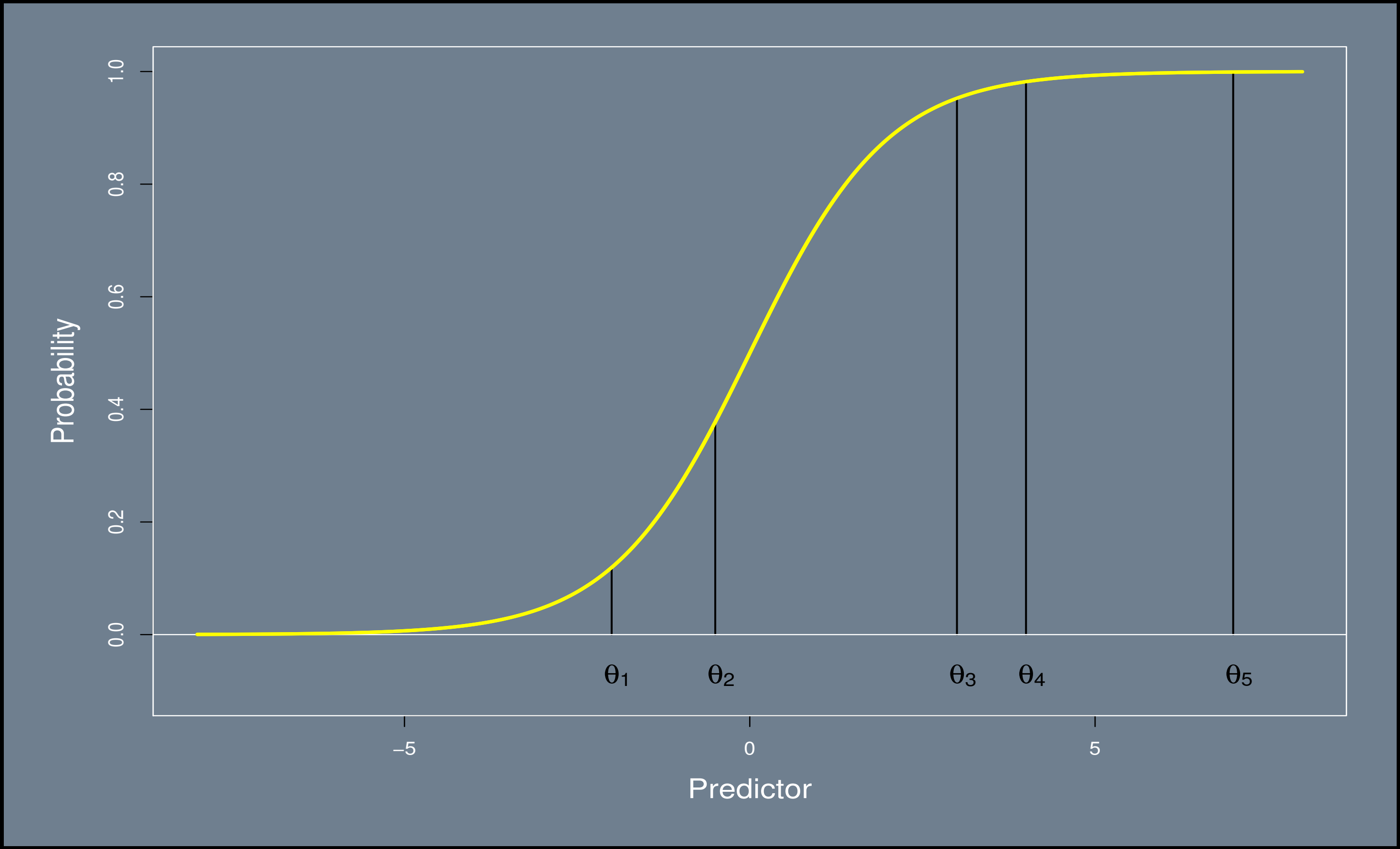
$$F_{\mathbf{E}}(\theta_r - \mathbf{X}'\boldsymbol{\beta}) = P(\mathbf{Y} \le r|\mathbf{X}) = [1 + \exp(-\theta_r + \mathbf{X}'\boldsymbol{\beta})]^{-1}$$

- The likelihood function is:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{X}, \mathbf{Y}) = \prod_{i=1}^{n} \prod_{j=1}^{C-1} [\Lambda(\theta_j - \mathbf{X}_i'\boldsymbol{\beta}) - \Lambda(\theta_{j-1} - \mathbf{X}_i'\boldsymbol{\beta})]^{z_{ij}}$$

  where $z_{ij} = 1$ if the ith case is in the jth category, and $z_{ij} = 0$ otherwise.

Threshold Illustration



Smoking Example from Fahrmeier and Tutz, First Edition, Page 90

```
library(MASS); library(nnet)

Freq <- c(577,164,192,145,682,245,27,4,20,15,46,47,7,0,3,7,11,27)
breathing.df <- data.frame(Freq, expand.grid(Age=1:2,Smoking.Status=1:3,
                                             Breathing.Status=1:3))

breathing.df$Age <- factor(breathing.df$Age)
levels(breathing.df$Age) <- c("< 40","40-59")

breathing.df$Smoking.Status <- factor(breathing.df$Smoking.Status)
levels(breathing.df$Smoking.Status) <- c("Never Smoked","Former Smoker",
                                         "Current Smoker")

breathing.df$Breathing.Status <- factor(breathing.df$Breathing.Status)
levels(breathing.df$Breathing.Status) <- c("Normal","Borderline","Abnormal")
breathing.df$Breathing.Status <- factor(as.ordered(breathing.df$Breathing.Status))
```

Smoking Data, `breathing.df`

```
   Freq   Age Smoking.Status Breathing.Status
1   577  < 40   Never Smoked           Normal
2   164 40-59   Never Smoked           Normal
3   192  < 40  Former Smoker           Normal
4   145 40-59  Former Smoker           Normal
5   682  < 40 Current Smoker           Normal
6   245 40-59 Current Smoker           Normal
7    27  < 40   Never Smoked       Borderline
8     4 40-59   Never Smoked       Borderline
9    20  < 40  Former Smoker       Borderline
10   15 40-59  Former Smoker       Borderline
11   46  < 40 Current Smoker       Borderline
12   47 40-59 Current Smoker       Borderline
13    7  < 40   Never Smoked         Abnormal
14    0 40-59   Never Smoked         Abnormal
15    3  < 40  Former Smoker         Abnormal
16    7 40-59  Former Smoker         Abnormal
17   11  < 40 Current Smoker         Abnormal
18   27 40-59 Current Smoker         Abnormal
```

Smoking Contrasts

- The sum contrast is "Effect Coding" and sums to zero down columns.

```
contrasts(breathing.df$Age) <- contr.sum(2)
contr.sum(2)
  [,1]
1    1
2   -1

contrasts(breathing.df$Smoking.Status) <- contr.sum(3)
contr.sum(3)
  [,1] [,2]
1    1    0
2    0    1
3   -1   -1
```

- Do not use the available `helmert` and `poly` as they are confusing ("Helmert coding compares each level of a categorical variable to the mean of the subsequent levels" from SAS documentation).

Smoking Contrasts

- The treatment contrast is more common, often called "Dummy Coding."

```
contr.treatment(2)
  2
1 0
2 1

contr.treatment(3)
  2 3
1 0 0
2 1 0
3 0 1

options()$contrasts
        unordered           ordered
"contr.treatment"      "contr.poly"

options("contrasts" = c("contr.treatment","contr.sum"))
```

Ordered Model Specification

```
breathing.plo <- polr(Breathing.Status ~ Age+Smoking.Status,data=breathing.df,
                      weights=Freq);    summary(breathing.plo)
Coefficients:
                 Value Std. Error t value
Age1            -0.389     0.0741   -5.24
Smoking.Status1 -0.581     0.1281   -4.53
Smoking.Status2  0.201     0.1255    1.60

Intercepts:
                    Value  Std. Error t value
Normal|Borderline    2.223  0.083      26.642
Borderline|Abnormal  3.685  0.143      25.694
```

```
contrasts(breathing.df$Age)
      [,1]
< 40     1
40-59   -1
```

```
contrasts(breathing.df$Smoking.Status)
               [,1] [,2]
Never Smoked      1    0
Former Smoker     0    1
Current Smoker   -1   -1
```

Interpretation

- Predicted probabilities need to use: $P(\mathbf{Y} \le r|\mathbf{X}) = [1 + \exp(-\theta_r + \mathbf{X}'\boldsymbol{\beta})]^{-1}$ after estimation of $\boldsymbol{\beta}$.
- Consider `Smoking.Status1 -0.581`, which gets the contrast `Never Smoked 1 0`, and being in the `Normal` category which has the upper threshold `Normal|Borderline 2.223`, plus being in the first age category.
- Thus we are interested in:

$$
\begin{aligned}
p(\mathbf{Y} \le 1|\mathbf{X}) = p(\mathbf{U} \le \theta_1) = p(\mathbf{U} \le 2.223) &= p(\mathbf{X}\boldsymbol{\beta} + \mathbf{E} \le 2.223) \\
&= p((1)(-0.581) + (0)(0.201) + \mathbf{X}_{Age1}\boldsymbol{\beta}_{Age1} + \mathbf{E} \le 2.223) \\
&= p(-0.581 + (1)(-0.389) + \mathbf{E} \le 2.223) \\
&= p(-0.970 + \mathbf{E} \le 2.223) \\
&= p(z \le 3.193)
\end{aligned}
$$

solved by `ilogit(3.193) [1] 0.9606`

Interpretation

- Compare with `Smoking.Status2 0.201`, which gets contrast `Former Smoker 0 1`, being in the `Normal` category plus being in the first age category.

$$p(\mathbf{Y} \leq 1|\mathbf{X}) = p((0))(-0.581) + (1)(0.201) + \mathbf{X}_{Age1}\boldsymbol{\beta}_{Age1} + \mathbf{E} \leq 2.223)$$

$$p(\mathbf{Y} \leq 1|\mathbf{X}) = p(0.201 - 0.389 + \mathbf{E} \leq 2.223)$$

$$p(\mathbf{Y} \leq 1|\mathbf{X}) = p(z \leq 2.411)$$

  solved by `ilogit(2.411) [1] 0.9177`

- So in the first case it is "easier" to be less than `2.223` with the negative contribution.
- For completeness, change to the the third smoking category which gets the contrast `Current Smoker -1 -1`:

$$p(\mathbf{Y} \leq 1|\mathbf{X}) = p((-1)(-0.581) + (-1)(0.201) + \mathbf{X}_{Age1}\boldsymbol{\beta}_{Age1} + \mathbf{E} \leq 2.223)$$

$$p(\mathbf{Y} \leq 1|\mathbf{X}) = p(0.581 - 0.201 - 0.389 + \mathbf{E} \leq 2.223)$$

$$p(\mathbf{Y} \leq 1|\mathbf{X}) = p(z \leq 2.232)$$

  solved by `ilogit(2.232) [1] 0.9031`

In Practice. . .

```
cbind(breathing.df, predict(breathing.plo,type="probs"))
1   577  < 40   Never Smoked          Normal 0.94446   0.042259 0.0132844
2   164 40-59   Never Smoked          Normal 0.97632   0.018157 0.0055213
3   192  < 40  Former Smoker          Normal 0.89437   0.079307 0.0263269
4   145 40-59  Former Smoker          Normal 0.86716   0.098960 0.0338815
5   682  < 40 Current Smoker          Normal 0.92317   0.058136 0.0186968
6   245 40-59 Current Smoker          Normal 0.76337   0.170371 0.0662618
7    27  < 40   Never Smoked      Borderline 0.94446   0.042259 0.0132844
8     4 40-59   Never Smoked      Borderline 0.97632   0.018157 0.0055213
9    20  < 40  Former Smoker      Borderline 0.89437   0.079307 0.0263269
10   15 40-59  Former Smoker      Borderline 0.86716   0.098960 0.0338815
11   46  < 40 Current Smoker      Borderline 0.92317   0.058136 0.0186968
12   47 40-59 Current Smoker      Borderline 0.76337   0.170371 0.0662618
13    7  < 40   Never Smoked        Abnormal 0.94446   0.042259 0.0132844
14    0 40-59   Never Smoked        Abnormal 0.97632   0.018157 0.0055213
15    3  < 40  Former Smoker        Abnormal 0.89437   0.079307 0.0263269
16    7 40-59  Former Smoker        Abnormal 0.86716   0.098960 0.0338815
17   11  < 40 Current Smoker        Abnormal 0.92317   0.058136 0.0186968
18   27 40-59 Current Smoker        Abnormal 0.76337   0.170371 0.0662618
```

Ordered Probit

- Suppose we specify instead that:

$$\epsilon \sim N(0,1)$$

- For the observed vector $\mathbf{Y}$:

$$\begin{aligned} p(\mathbf{Y} \leq r|\mathbf{X}) &= p(\mathbf{U} \leq \theta_r) \\ &= p(\mathbf{X}\boldsymbol{\beta} + \mathbf{E} \leq \theta_r) \\ &= p(\mathbf{E} \leq \theta_r - \mathbf{X}\boldsymbol{\beta}) \\ &= F_{\mathbf{E}}(\theta_r - \mathbf{X}\boldsymbol{\beta}) \\ &= \Phi(\theta_r - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

Ordered Probit

- So the individual probabilities are:

$$
\begin{aligned}
p(Y \leq 1|\mathbf{X}) &= p(Y = 1) = \Phi(\theta_1 - \mathbf{X}\boldsymbol{\beta}) \\
p(Y \leq 2|\mathbf{X}) &= p(Y = 1) + p(Y = 2) = \Phi(\theta_2 - \mathbf{X}\boldsymbol{\beta}) \\
p(Y \leq 3|\mathbf{X}) &= p(Y = 1) + p(Y = 2) + p(Y = 3) = \Phi(\theta_3 - \mathbf{X}\boldsymbol{\beta}) \\
&\vdots \\
p(Y \leq k-1|\mathbf{X}) &= 1 - p(Y = k) = \Phi(\theta_{k-1} - \mathbf{X}\boldsymbol{\beta}) \\
p(Y \leq k|\mathbf{X}) &= 1
\end{aligned}
$$

Ordered Probit

- We can also look at these as differences:

$$
\begin{aligned}
p(Y = 1|\mathbf{X}) &= \Phi(\theta_1 - \mathbf{X}\boldsymbol{\beta}) \\
p(Y = 2|\mathbf{X}) &= \Phi(\theta_2 - \mathbf{X}\boldsymbol{\beta}) - \Phi(\theta_1 - \mathbf{X}\boldsymbol{\beta}) \\
p(Y = 3|\mathbf{X}) &= \Phi(\theta_3 - \mathbf{X}\boldsymbol{\beta}) - \Phi(\theta_2 - \mathbf{X}\boldsymbol{\beta}) \\
&\vdots \\
p(Y = k|\mathbf{X}) &= \Phi(\theta_k - \mathbf{X}\boldsymbol{\beta}) - \Phi(\theta_{k-1} - \mathbf{X}\boldsymbol{\beta}) \\
&= 1 - \Phi(\theta_{k-1} - \mathbf{X}\boldsymbol{\beta})
\end{aligned}
$$

Contraception in El Salvador

- A classic sociological and public policy question.
- Report of the Demographic and Health Survey conducted in El Salvador in 1985 (FESAL-1985).
- Data: current Use of Contraception By Age, Currently Married Women. El Salvador, 1985.
- 3165 currently married women classified by age, grouped in five-year intervals, and current use of contraception, classified as: sterilization, other methods, and no method.
- Data setup:

```
library(MASS);   library(nnet)
contraception.mat <-
  as.matrix(read.table(
  "https://jeffgill.org/wp-content/uploads/2024/08/contraception.dat_.txt",
  header=TRUE))
contraception.df <- data.frame(expand.grid(Response=1:3,
                        "Age"=contraception.mat[,1]),
                        "Freq"=as.numeric(t(contraception.mat[,2:4])))
```

Contraception in El Salvador, Data Setup

```
contraception.df$Response<- factor(contraception.df$Response)
        levels(contraception.df$Response) <- c("Sterilization","Other","None")
contraception.df$Age<- factor(contraception.df$Age)
        levels(contraception.df$Age) <-
               c("15-19","20-24","25-29","30-34","35-39","40-44","45-49")

contraception.mat            Age Sterilization Other None All
                        [1,]  15             3    61  232 296
                        [2,]  20            80   137  400 617
                        [3,]  25           216   131  301 648
                        [4,]  30           268    76  203 547
                        [5,]  35           197    50  188 435
                        [6,]  40           150    24  164 338
                        [7,]  45            91    10  183 284
```

Contraception in El Salvador, Ordered Logit Model

```
contraception.plo <- polr(Response ~ Age, weights=Freq,data=contraception.df)
summary(contraception.plo)

Coefficients:
              Value Std. Error    t value
Age20-24 -0.6919165  0.1605822  -4.308801
Age25-29 -1.5057925  0.1569331  -9.595125
Age30-34 -2.0247671  0.1614506 -12.541096
Age35-39 -1.8161068  0.1668556 -10.884303
Age40-44 -1.6773177  0.1752633  -9.570273
Age45-49 -0.9936089  0.1861528  -5.337597

Intercepts:
                    Value    Std. Error t value
Sterilization|Other  -2.1252   0.1410   -15.0697
Other|None           -1.4170   0.1386   -10.2262

Residual Deviance: 5963.335
AIC: 5979.335
```

Contraception in El Salvador, Ordered Probit Model

```
contraception.pro <- polr(Response ~ Age, weights=Freq,data=contraception.df,
    method = c("probit"))
summary(contraception.pro)

Coefficients:
              Value Std. Error    t value
Age20-24 -0.4476988 0.09466000  -4.729546
Age25-29 -0.9715711 0.09281219 -10.468141
Age30-34 -1.2936516 0.09530248 -13.574165
Age35-39 -1.1657064 0.09869954 -11.810656
Age40-44 -1.0829845 0.10363113 -10.450378
Age45-49 -0.6860022 0.10911112  -6.287189

Intercepts:
                    Value    Std. Error t value
Sterilization|Other  -1.3569   0.0820   -16.5502
Other|None           -0.9201   0.0807   -11.4070

Residual Deviance: 5944.509             AIC: 5960.509
```

Questions From These Results

- Comparing the BIC for these two models:

  The log likelihoods are most easily calculated by:

$$\ell()_{polr} = p - \frac{1}{2}\text{AIC}_{polr} = 6 - \frac{1}{2}(5979.34) = -2983.670$$
$$\ell()_{popr} = p - \frac{1}{2}\text{AIC}_{popr} = 6 - \frac{1}{2}(5960.51) = -2974.255$$

  The BIC is then calculated by:

$$\text{BIC}_{polr} = -2\ell()_{polr} + p\log(n) = -2 \times -2983.670 + 6 \times \log(3165) = 6015.70$$
$$\text{BIC}_{popr} = -2\ell()_{popr} + p\log(n) = -2 \times -2974.255 + 6 \times \log(3165) = 5996.87$$

Extensions in R

- The package **mlogit** can handle heteroscedastic, nested and random parameter models.

- The package **ordinal** accomodates multiple random effect terms and they may be nested, crossed or partially nested/crossed. Restrictions of symmetry and equidistance can be imposed on the thresholds.

- The package **oglmix** provides ordered logit and probit where the error variance does not have to be constant across observations by allowing a variance distribution instead.

```
install.packages("oglmix"); library(oglmix)
breathing.het <- oglmix(Breathing.Status~Age+Smoking.Status+Age:Smoking.Status,
                        data=breathing.df, link="logit", constantMEAN=FALSE,
            constantSD=FALSE, threshparam=NULL)
```

- The package **RSGHB** does Hierarchical Bayesian modeling of ordinal outcomes. The large class of supported modeles includes ordered probit, ordered logit as well as multinomial logit, mixed logit, nested logit, error components logit, and latent class models. Parameters can be fixed or random in the specifications.

- The package **Rchoice** does ordered probit and logit (and Poisson) with random parameters for cross-sectional and longitudinal data.

Pediatric Neurocritcal Care

- Pineda *etal.*, Lancet–Neurology 2013, "Effect of Implementation of a Paediatric Neurocritical Care Programme On Outcomes After Severe Traumatic Brain Injury: A Cohort Study."
- 10 years of PTBI data with a change in the middle-point (September 2005).
- PNCP (Pediatric Neurocritical Care Program): a time-sensitive, severity-based approach to monitor and treat children with TBI that coordinated communication and activity amongst PICU staff and physician faculty and trainees, conforming with the 2003 Brain Trauma Foundation guidelines.
- This included a detailed training program, an explicit process for maintaining pathway fidelity, and continuous quality improvement.
- Groups: $n_{Pre-PNCP} = 63$, $n_{Post-PNCP} = 60$, treated as a fixed effect variable (treatment contrast).
- Tests for differences in demographics between the two periods failed to find statistically reliable differences.
- Outcomes: Medical Examiner/Morgue, Different Acute Care Hospital, Inpatient Rehab Facility, Home With Healthcare, Home With Outpatient Rehab, Home With No Assistance.

Results from the Ordered Probit Model

	Coefficient	Std.Err.	t-value
Post-PNCP	0.482477	0.216061	2.233
Age In Months	-0.004674	0.002127	-2.198
White	-0.318926	0.129315	-2.466
Length of Stay in PICU	-0.003776	0.007839	-0.482
Male	0.111984	0.107548	1.041
ICP Monitoring	0.997479	0.299579	3.330
Post-Resuscitation GCS	0.125677	0.060159	2.089
PRISM III	-0.065137	0.018125	-3.594
Injury Severity Score^2	-0.000315	0.000134	-2.345
Fall	0.291087	0.268258	1.085
Motor Vehicle Accident	0.197797	0.191271	1.034
Pedestrian Accident	0.147976	0.241442	0.613

NOTES:

- Reference category for the injury etiologies is "Other."
- Race (white) `-0.318926`, means that moving from 0=non-white to 1=white pushed the expected outcome down the scale of U towards more unfavorable outcomes.
- Coefficients such as intracranial pressure Monitoring `0.997479`, have the opposite effect.

Ordered Probit Threshold Estimates

Threshold	Categories Separated	Coefficient	Std. Error	t-value
θ_1	Medical Examiner/Morgue *to* Different Acute Care Hospital	-0.647	0.188	-3.442
θ_2	Different Acute Care Hospital *to* Inpatient Rehab Facility	-0.377	0.226	-1.670
θ_3	Inpatient Rehab Facility *to* Home With Healthcare	0.979	0.262	3.733
θ_4	Home With Healthcare *to* Home With Outpatient Rehab	1.005	0.262	3.829
θ_5	Home With Outpatient Rehab *to* Home With No Assistance	1.433	0.266	5.391

NOTES:

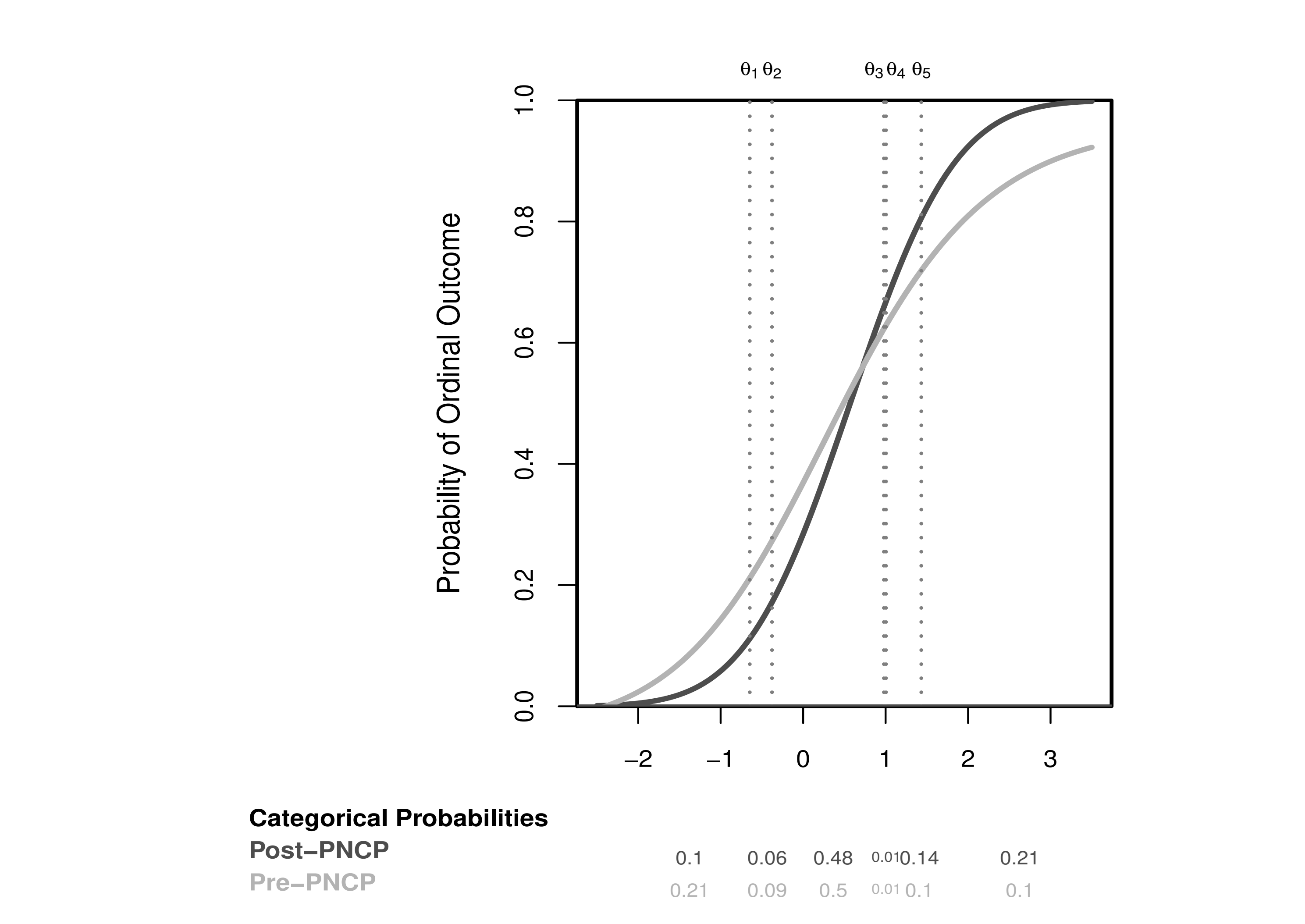
- The literal value of these coefficients is unimportant.
- The statistical significance of these coefficients is unimportant.
- They are important only to "help" estimate the β coefficients.

Making a Prediction Difference Using Race

- Suppose a predicted outcome on the U metric for a particular white patient was -0.4 (given values for all of the other $\mathbf{X}$ variables).
- Then the model would predict the category [Different Acute Care Hospital].
- If this was a non-white patient, the coefficient β_{Race} would assign -0.318926 instead of 0 in the $\mathbf{X}_i\boldsymbol{\beta}$ calculation.
- This reduction gives the (hypothetical) patient a prediction of $U_i = -0.718926$, which corresponds to the categorical prediction of [Medical Examiner/Morgue].

Graphical Comparison

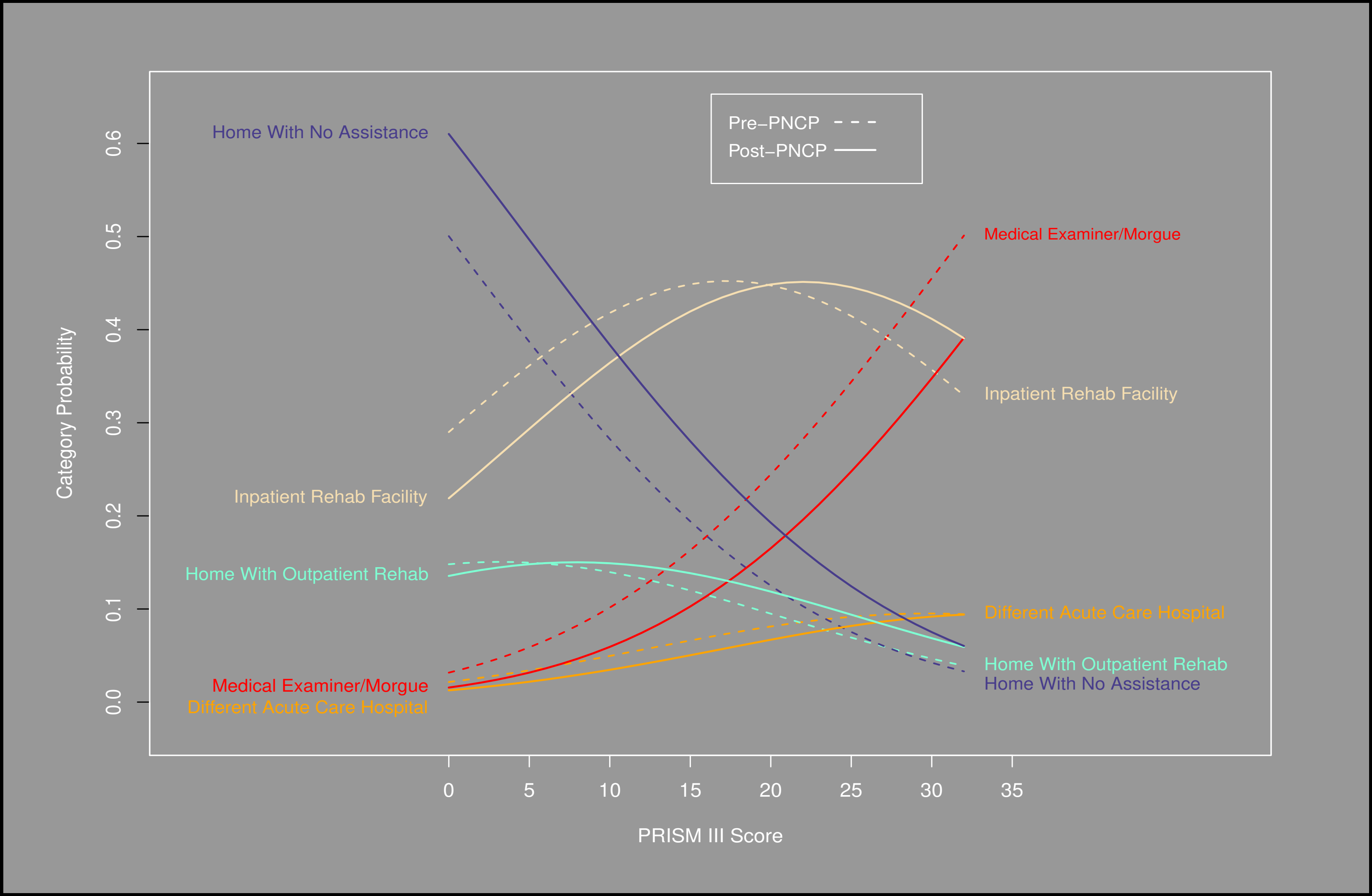
- The x-axis is the U metric and the y-axis is probability.
- The five θ cutpoints are given by the dotted vertical lines and labeled at the top.
- The slices give the probability for being in each of the categories for a white male following a motor vehicle accident, with all other explanatory variables set at the data mean.



First Differences for This Model

- Select two levels of one explanatory variable and setting all others at their means: $\mathbf{X}_1$ and $\mathbf{X}_2$.
- Two hypothetical probability vectors are created by applying the link function to $\mathbf{X}_1\boldsymbol{\beta}$ and $\mathbf{X}_2\boldsymbol{\beta}$, which can be compared.
- Race is set at white, etiology is set at motor vehicle accident, and the two vectors are then made different with the Group variable: one indicates Pre-PNCP and the other indicates Post-PNCP status.
- All other variables except Group, Race and Etiology are set at the data means.
- These almost identical cases are multiplied by the estimated regression coefficient vector and the ordered probit link function then transforms each onto the probability scale for comparison.
- The probability of death falls 53% following PNCP initiation (from 0.210 to 00988) and the probability for discharge to home with no assistance increases 53% (from 0.101to 0.214).

Smooth Predictions From the Model



Unordered Background

- The multinomial distribution is an extension of the binomial where the outcome is allowed to take on more than two values.
- Define Y_i as the *nominal* random variable taking on values $1, 2, \ldots, J$.
- Let $p_{ij} = p(Y_i = j)$ with the requirement that $\sum_{j=1}^{J} p_{ij} = 1$.
- Further define Y_{ij} as the number of observations falling into outcome j for case i.
- For *Grouped Data* types, these are cell counts where $n_i = \sum_j Y_{ij}$.
- For *Ungrouped Data* types, we have the restriction that $n_i = 1$ for exactly one outcome and $n_i = 0$ for the rest.
- The PMF is then given by:

$$p(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{ij}) = \frac{n_i}{y_{i1}! \cdots y_{iJ}} p_{i1}^{y_{i1}} \cdots p_{iJ}^{y_{iJ}}$$

What is IIA?

- Independence from Irrelevant Alternatives.
- This is equivalent to the reasonable assumption of iid errors from the model.
- A person's 's probability of selecting one choice alternative over another is not affected by the presence or absence of a third alternative.
- Put another way, any item added to the set of choices will decrease all other items' probability by an equal fraction.
- More technically, the odds ratio between any two choices does not depend on the other choices.
- Simple test for IIA (McFadden 1976): remove each choice one-at-a-time, re-run the model, and check to see if coefficients differ considerably.
- More formal test given in Hausman (1978), Hausman & McFadden (1984), and Small & Hsiao (1985).

Example of IIA

- We have two candidates during a debate and are indifferent indifferent:

$$p(\text{Democrat}) = p(\text{Republican in blue suit}) = \frac{1}{2}.$$

- So the odds of voting for the Democrat over the Republican in the blue suite are:

$$\frac{p(\text{Democrat})}{p(\text{Republican in blue suit})} = 1.$$

- Suppose we add a choice of a Republican in a black suite such that we are still indifferent between choices:

$$p(\text{Democrat}) = p(\text{Republican in blue suit}) = p(\text{Republican in black suit}) = \frac{1}{3},$$

  which is the IIA assumption since $p(\text{Democrat})/p(\text{Republican in blue suit}) = 1$.

- However, this is at odds with the realistic notion that people should be indifferent between Republicans:

$$p(\text{Republican in blue suit}) = p(\text{Republican in black suit}) = \frac{1}{4}.$$

  since now:

$$\frac{p(\text{Democrat})}{p(\text{Republican in blue suit})} = 2.$$

What is IIA? (real example)

- Suppose a country has a liberal and conservative party, and a new conservative party enters.
- IIA imples that the entrance of the 2nd conservative party should not affect the relative probability of an individual choosing between the liberal party and the first conservative party.
- Hypothesized real example of violating IIA, French presidential election of 2002:

	April 21		May 5	
Jacques Chirac	5,665,855	19.88%	25,537,956	82.21%
Jean-Marie Le Pen	4,804,713	16.86%	5,525,032	17.79%
Lionel Jospin	4,610,113	16.18%		

- The idea was that some Chirac supporters voted for Le Pen to knock-out Jospin who would have been more competitive that Le Pen in the run-off.

A Quick Test

- Define:

$\hat{\boldsymbol{\beta}}_F$	coefficient vector under full set of outcome alternatives
$\hat{\boldsymbol{\beta}}_S$	coefficient vector under subset of outcome alternatives
$\hat{\Sigma}_F$	variance/covariance matrix under full
$\hat{\Sigma}_S$	variance/covariance matrix under subset
k	the length of $\hat{\boldsymbol{\beta}}_F$ and $\hat{\boldsymbol{\beta}}_S$.

  Note that $\hat{\boldsymbol{\beta}}_F$ and $\hat{\boldsymbol{\beta}}_S$ are of equal length since we are only changing the choice alternatives.

- Then compute the statistic:

$$\chi^2_k = \left(\hat{\boldsymbol{\beta}}_S - \hat{\boldsymbol{\beta}}_F\right)' \left[\hat{\Sigma}_S - \hat{\Sigma}_F\right]^{-1} \left(\hat{\boldsymbol{\beta}}_S - \hat{\boldsymbol{\beta}}_F\right)$$

  where tail values imply a difference and non-tail values mean that you cannot reject the null hypothesis of IIA. where tail values indicate a problem.

- This is a low power test (power is the probability of rejecting a false null).

- More on IIA tests later.

Multinomial Logit

- Sometimes also called the *multiple logit model*.

- Applications in political science research: Abramson, *et al.* 1992; Canache, Mondak, & Conroy 1994; Gerber 1996; Iversen 1994; Layman & Carmines 1997; Powers & Cox 1997; Quinn, Martin, & Whitford 1999; Wahlbeck 1997; Martinez & Gill 2005.

- The resulting coefficient-sets (one for each $J-1$ choices distinct from the reference category) provide the relative effect through the logit function of that explanatory variable on the probability that the respondent chose category j rather than this reference category.

- J different parameter *vectors* $\boldsymbol{\beta}_j, \quad j = 1 \ldots J$, the first of which is all zeros, $\boldsymbol{\beta}_1 = \mathbf{0}$ representing the reference category.

- Warning: this does *not* mean that $p_{i1} = 0$.

- In finite samples, it is standard to assume that the error matrix is multivariate Weibull. The non-multivariate Weibull PDF looks like this: $f(x|\gamma,\beta) = \frac{\gamma}{\beta}x^{\gamma-1}\exp(-x^{\gamma}/\beta)$ for $x \geq 0, \gamma, \beta > 0$.

Multinomial Logit

- The probability that respondent i chooses category j over category 1 is given by:

$$p_{ij} = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{k=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_k)}$$

- How does this compare to the regular logit we've come to know? Suppose that $J = 2$ above, then there would be $\boldsymbol{\beta}_1 = \mathbf{0}$ and $\boldsymbol{\beta}_2$ estimated:

$$p_{i2} = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_2)}{\sum_{k=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_k)} = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_2)}{\exp(\mathbf{X}_i \boldsymbol{\beta}_1) + \exp(\mathbf{X}_i \boldsymbol{\beta}_2)} = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_2)}{\exp(0) + \exp(\mathbf{X}_i \boldsymbol{\beta}_2)} = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_2)}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta}_2)}$$

- The log-likelihood function is:

$$\ell(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J | \mathbf{X}) = \sum_{j=1}^{J} \sum_{y_i = j} \exp(\mathbf{X}_i \boldsymbol{\beta}_j) - \sum_{l=1}^{n} \log\left(1 + \sum_{j=1}^{J} \exp(\mathbf{X}_\ell \boldsymbol{\beta}_j)\right),$$

  where $\sum_{y_i=j}$ sums over cases where the outcome variable is the jth category.

- This is globally concave and the routine is canned in virtually all of the user-friendly software packages including `R`.

Multinomial Logit

- Assumes that IIA holds.
- This is because:

$$\frac{p_j}{p_\ell} = \frac{\exp(\mathbf{X}\boldsymbol{\beta}_j)}{\sum_{k=1}^{J}\exp(\mathbf{X}\boldsymbol{\beta}_k)} \times \left[\frac{\exp(\mathbf{X}\boldsymbol{\beta}_\ell)}{\sum_{k=1}^{J}\exp(\mathbf{X}\boldsymbol{\beta}_k)}\right]^{-1} = \frac{\exp(\mathbf{X}\boldsymbol{\beta}_j)}{\exp(\mathbf{X}\boldsymbol{\beta}_\ell)}.$$

- Therefore in log terms:

$$\log\left(\frac{p_j}{p_\ell}\right) = (\boldsymbol{\beta}_j - \boldsymbol{\beta}_\ell)\mathbf{X}$$

- This means that the odds ratio between any two choices does not include information from another choice, *so it cannot accomodate information from an added choice.*

MNL Estimation

- Understanding MNL coefficient estimates:

$$\log\left[\frac{p_{ij}}{p_{i1}}\right] = \log\left[\frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_j)/\sum_{k=1}^{J}\exp(\mathbf{X}_i\boldsymbol{\beta}_k)}{\exp(\mathbf{X}_i\boldsymbol{\beta}_1)/\sum_{k=1}^{J}\exp(\mathbf{X}_i\boldsymbol{\beta}_k)}\right] = \log\left[\frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_j)}{\exp(\mathbf{X}_i 0)}\right] = \log\left[\frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_j)}{1}\right] = \mathbf{X}_i\boldsymbol{\beta}_j,$$

  the log of the ratio of probability of selecting choice j to the probability of selecting the baseline choice.

- Note that this uses the baseline assumption $\boldsymbol{\beta}_1 = \mathbf{0}$ for the the reference category.

- It should be intuitive that the sum of the probabilities equals one for every respondent since this covers all choice alternatives:

$$1 = \sum_{k=1}^{J} p_{ij}$$

MNL Estimation

- Using the sum property we can solve for any of the the individual choice probabilities.
- To get the probability of the reference category start with the log-odds:

$$\log\left[\frac{p_{ij}}{p_{i1}}\right] = \mathbf{X}_i\boldsymbol{\beta}_j \longrightarrow p_{ij} = \exp(\mathbf{X}_i\boldsymbol{\beta}_j)p_{i1}.$$

and use the total probability property:

$$p_{i1} = 1 - \sum_{j=2}^{J} p_{ij} = 1 - \sum_{j=2}^{J} \exp(\mathbf{X}_i\boldsymbol{\beta}_j)p_{i1} = 1 - p_{i1}\sum_{j=2}^{J} \exp(\mathbf{X}_i\boldsymbol{\beta}_j)$$

$$1 = \frac{1}{p_{i1}} - \sum_{j=2}^{J} \exp(\mathbf{X}_i\boldsymbol{\beta}_j)$$

$$p_{i1} = \left(1 + \sum_{j=2}^{J} \exp[\mathbf{X}_i\boldsymbol{\beta}_j]\right)^{-1}.$$

Faraway's NES Analysis

```
# LOAD LIBRARY AND DATA, FIX ANES DATA
library(faraway)
data(nes96)
sPID <-  nes96$PID
levels(sPID) <- c("Democrat","Democrat","Independent","Independent","Independent",
                  "Republican", "Republican")
summary(sPID)

   Democrat Independent  Republican
        380         239         325

inca <- c(1.5,4,6,8,9.5,10.5,11.5,12.5,13.5,14.5,16,18.5,21,23.5,27.5,32.5,37.5,
          42.5,47.5,55,67.5,82.5,97.5,115)
nincome <- inca[unclass(nes96$income)]
```

Faraway's NES Analysis

```
summary(nincome)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.5    23.5    37.5    46.6    67.5   115.0
table(nes96$educ)

    MS HSdrop     HS   Coll  CCdeg  BAdeg  MAdeg
    13     52    248    187     90    227    127

cutinc <- cut(nincome,7)
il <- c(8,26,42,58,74,90,107)
cutage <- cut(nes96$age,7)
al <- c(24,34,44,54,65,75,85)
```

Faraway's NES Analysis

```
library(nnet)

# RUN MODEL WITH ANES DATA
mmod <- multinom(sPID ~ age + educ + nincome, nes96)
summary(mmod)

Coefficients:
            (Intercept)      age educ.L educ.Q educ.C  educ^4 educ^5  educ^6 nincome
Independent       -1.20 0.000153 0.0635 -0.122  0.112 -0.0766  0.136  0.1543  0.0162
Republican        -1.64 0.008194 1.1941 -1.229  0.154 -0.0283 -0.122 -0.0374  0.0172

Std. Errors:
            (Intercept)     age educ.L educ.Q educ.C educ^4 educ^5 educ^6 nincome
Independent       0.327 0.00537  0.457  0.414  0.350  0.288  0.249  0.217 0.00311
Republican        0.331 0.00490  0.650  0.604  0.487  0.361  0.270  0.203 0.00288

Residual Deviance: 1968.3
AIC: 2004.3
```

Faraway's NES Analysis

```
N <- factor(Nlevs <- c(1,4,8,22,40,51,65));   contr.poly(N)
                .L            .Q            .C         ^4            ^5          ^6
[1,] -5.669467e-01  5.455447e-01 -4.082483e-01  0.2417469 -1.091089e-01  0.03289758
[2,] -3.779645e-01  1.133705e-16  4.082483e-01 -0.5640761  4.364358e-01 -0.19738551
[3,] -1.889822e-01 -3.273268e-01  4.082483e-01  0.0805823 -5.455447e-01  0.49346377
[4,]  2.855301e-17 -4.364358e-01  5.779362e-17  0.4834938 -1.231475e-15 -0.65795169
[5,]  1.889822e-01 -3.273268e-01 -4.082483e-01  0.0805823  5.455447e-01  0.49346377
[6,]  3.779645e-01 -5.621884e-17 -4.082483e-01 -0.5640761 -4.364358e-01 -0.19738551
[7,]  5.669467e-01  5.455447e-01  4.082483e-01  0.2417469  1.091089e-01  0.03289758
```

```
contr.sum(N)
   [,1] [,2] [,3] [,4] [,5] [,6]
1     1    0    0    0    0    0
4     0    1    0    0    0    0
8     0    0    1    0    0    0
22    0    0    0    1    0    0
40    0    0    0    0    1    0
51    0    0    0    0    0    1
65   -1   -1   -1   -1   -1   -1
```

NES RE-Analysis

```
options("contrasts" = c("contr.treatment","contr.sum"))
mmod <- multinom(sPID ~ age + educ + nincome, nes96)
summary(mmod)

Coefficients:
        (Intercept)      age  educ1   educ2  educ3   educ4 educ5   educ6 nincome
Independent   -1.20 0.000154 -0.176  0.0938 0.0693 -0.0854 0.150 -0.0683  0.0162
Republican    -1.64 0.008195 -1.406 -0.4182 0.2857  0.5475 0.478  0.4650  0.0172

Std. Errors:
            (Intercept)     age educ1 educ2 educ3 educ4 educ5 educ6 nincome
Independent       0.327 0.00537 0.591 0.312 0.186 0.212 0.264 0.205 0.00311
Republican        0.331 0.00490 0.915 0.373 0.215 0.225 0.276 0.224 0.00288

Residual Deviance: 1968.333             AIC: 2004.333

mmod.null <- multinom(sPID ~ 1,data=nes96)
Residual Deviance: 2041
```

Faraway Doing Stepwise Analysis (Don't Do This!)

```
mmodi <- step(mmod);   summary(mmodi)                    # SHOWING ONLY THE LAST STEP...
Coefficients:                                       Std. Errors:
            (Intercept)  nincome                                 (Intercept)   nincome
Independent    -1.17493 0.016087                    Independent      0.15361 0.0028497
Republican     -0.95036 0.017665                    Republican       0.14169 0.0026525

Residual Deviance: 1985.4
AIC: 1993.4

# NOW DROP EDUCATION FROM THE RHS
mmode <- multinom(sPID ~ age +  nincome, nes96)

mmod$edf                mmode$edf
[1] 18                  [1] 6

deviance(mmode) - deviance(mmod)                             [1] 16.206

pchisq(16.206,mmod$edf-mmode$edf,lower=F)                    [1] 0.18198
```

Faraway's NES Analysis

```
# LOOK AT PREDICTIONS FOR LEVELS OF il
predict(mmodi,data.frame(nincome=il),type="probs")

  Democrat Independent Republican
1  0.55663     0.19552    0.24786
2  0.48049     0.22546    0.29405
3  0.41343     0.25094    0.33564
4  0.34939     0.27432    0.37629
5  0.29033     0.29486    0.41481
6  0.23758     0.31211    0.45031
7  0.18917     0.32668    0.48415

predict(mmodi,data.frame(nincome=il))   # GIVES MOST PROBABLE CATEGORY

[1] Democrat   Democrat   Democrat   Republican Republican Republican Republican
Levels: Democrat Independent Republican
```

Faraway's NES Analysis

```
# TWO WAYS TO PREDICT FOR AN INCOME OF ZERO: STEPWISE RESULTS (mmodi model)
cc <- c(0,-1.17493,-0.95036)     # mmodi COEFFICIENTS EXCEPT nincome
cc <- c(0,summary(mmodi)$coefficients[,1])
exp(cc)/sum(exp(cc))
            Independent  Republican
    0.58982     0.18216     0.22802

predict(mmodi,data.frame(nincome=0),type="probs")

   Democrat Independent  Republican
    0.58982     0.18216     0.22802
```

Faraway's NES Analysis

```
# SLOPE TERMS ARE THE LOG-ODDS OF MOVING FROM BASELINE CATEGORY TO OTHERS FOR A
# 1-UNIT ($1000) CHANGE IN INCOME, DEM-INDEP THEN DEM-REP.
# STEPWISE RESULTS (mmodi model)
(pp <- predict(mmodi,data.frame(nincome=c(0,1)),type="probs"))

  Democrat Independent Republican
1  0.58982     0.18216    0.22802
2  0.58571     0.18382    0.23047

log(pp[1,1]*pp[2,2]/(pp[1,2]*pp[2,1]))

[1] 0.016087

log(pp[1,1]*pp[2,3]/(pp[1,3]*pp[2,1]))

[1] 0.017665

# RECOVERS THE COEFFICENTS FROM summary()
```

Trading Butter for Guns

DOMESTIC IMPERATIVES FOR FOREIGN POLICY SUBSTITUTION

DAVID H. CLARK
Department of Political Science
Binghamton University (SUNY)

The international relations literature largely presumes that leaders engage in foreign policy substitution but does not provide a compelling theoretical explanation or convincing empirical evidence that substitution occurs. This article offers a theory of foreign policy choice based on the differences between private and public goods. It assumes that private goods and public goods are useful under different circumstances and conditions. Leaders select a policy based on political needs, so private- and public-goods approaches are employed alternatively depending on domestic situations: policies are substituted one for another. The trade-off between aggressive unilateral economic behavior and military conflict as the United States conducted foreign policy during the cold war is examined. Results show that leaders facing economic concerns and/or domestic opposition prefer trade aggression, a patently private-good-like policy, and substitute such policies in response to changing domestic stimuli.

The vast array of policy options available to political leaders as they seek to accomplish substantive goals (enact policy) and achieve personal goals (retain office) prompts students of politics to theorize why leaders choose the policies they do. The essence of this question appears in international relations research positing that leaders substitute one foreign policy for another depending on the particular conditions they encounter at any given time (e.g., Most and Starr 1989; Regan 2000; Bennett and Nordstrom 2000; Morgan and Palmer 2000; Enterline and Gleditsch 2000). Substitution models often deal with specific political contexts rather than offering general explanations of what types of policies leaders are likely to prefer. In an implicit manner, diversionary use of force research suggests that domestically troubled leaders substitute force for action directed at correcting the source of the domestic trouble. Presumably, leaders are motivated to divert attention because they believe they lack the policy tools to correct the domestic problems they face. However, this argument and others that contend that domestic forces can increase the incentives for international conflict provide only a limited context within which leaders select policies. In reality, political leaders have at their disposal a large set of policies from which to choose; the

AUTHOR'S NOTE: Thanks for comments and suggestions go to Mark Crescenzi, Ray Dacey, Andrew Enterline, Kristian Gleditsch, Paul Huth, Doug Lemke, Will Moore, Bill Reed, Dale Smith, and Barclay Ward. An earlier version of the project was presented at the annual meeting of the American Political Science Association, Atlanta, Georgia, September 1999.

JOURNAL OF CONFLICT RESOLUTION, Vol. 45 No. 5, October 2001 636-660

- 0 = no conflict,
- 1 = military conflict,
- 2 = trade conflict, and
- 3 = both military and trade conflict.

Multinomial logit compares the reference category (0, no conflict) to the other categories and produces coefficients for all independent variables for each of the other three outcomes.[17] So, the results will indicate the effects of the covariates on the probability of a change from

- no conflict to military conflict,
- no conflict to trade conflict, and
- no conflict to both military and trade conflict.

The following section presents the results of these analyses and discusses the implications of the results for arguments about foreign policy substitution.

RESULTS AND DISCUSSION

The probit analyses, presented in two separate specifications in Table 1, provide strong initial support for the congruence hypothesis and the hypothesis regarding unemployment. Generally, they support the idea that U.S. presidents employ different tools depending on the domestic political and economic conditions they face, specifically that they select private-good-like solutions to deal with private-good-like problems.

The probit analysis in model 1 indicates a significant relationship between the level of presidential support in the Congress and the likelihood the United States will select to use military force rather than a GATT action. In fact, the impact of an increase in presidential support on the likelihood of military action is substantial: a 5% increase in support for the president results in a 4% increase in the likelihood that the United States will pursue military rather than economic action.[18]

Similarly, in model 2, the effect of unified government is to enable presidential military action. Institutional congruence increases the likelihood the United States will

17. The categories of this nominal dependent variable are distributed as

- 0 = no action = 61.13% of 600 monthly observations;
- 1 = MID only = 22.6% of 600 monthly observations;
- 2 = GATT only = 11.8% of 600 monthly observations;
- 3 = both MID and GATT = 4.3% of 600 monthly observations.

18. The effects of variables in probit models cannot be interpreted in the straightforward manner to which least squares models are amenable. Rather, marginal effects are computed by

$$\phi[\Sigma(\beta' X) + x_i\sigma] - \phi[\Sigma(\beta' X)],$$

or the change in predicted probability given a one standard deviation change in the variable of interest, other variables held constant at their means or modes. In the case of dichotomous independent variables, the effect reflects the change in that variable from 0 to 1 (modal to nonmodal value), others held constant.

TABLE 2
Multinomial Logit Models of U.S. Foreign Policy Options

Variable	$\hat{\beta}$	SE
Characteristics of Prob[$Y = 1$]: MID vs. no action		
Presidential support	0.012	0.009
Δ unemployment	–0.789*	0.577
Election year	0.147	0.280
Constant	–1.72**	0.821
Characteristics of Prob[$Y = 2$]: GATT vs. no action		
Presidential support	–0.044***	0.012
Δ unemployment	0.475	0.708
Election year	0.217	0.335
Constant	1.32*	0.821
Characteristics of Prob[$Y = 3$]: both vs. no action		
Presidential support	–0.032**	0.017
Δ unemployment	–0.283	1.07
Election year	–0.609	0.586
Constant	–0.133	1.196
Likelihood Ratio Tests	$-2LL \sim \chi^2$	
vs. null model	301.1***	
vs. excluding presidential support	56.36***	
vs. excluding Δ unemployment	117.79***	

NOTE: $N = 431$. Dependent variable indicates no conflict (0), the presence of a militarized dispute (1), a trade dispute (2), or both military and trade conflict simultaneously (3). GATT = General Agreement on Tariffs and Trade, MID = Militarized Industrial Dispute; –2LL evaluates the full model in comparison to the null model.
*$p \leq .10$. **$p \leq .05$. ***$p \leq .01$, one-tailed tests.

choose from among the four options identified above: no action, military action, trade action, and military and trade action at the same time.

Table 2 reports the multinomial logit results. However, a comment on statistical inference in these models is necessary prior to a discussion of the results. Because these models estimate the effects of independent variables on each category compared to the base category (in this case, no conflict), the models produce $j - 1$ (in this case, 3) parameter estimates for each independent variable. It is not entirely clear how to treat variables that produce some significant parameters and some insignificant parameters. It appears the most common solution is to conduct block log-likelihood tests on each variable to determine if each variable's inclusion significantly improves the model (see Greene 1997). I follow this convention and report block test results in Table 2.[22]

Discussion of multinomial logit coefficients is complicated not only by the inference problem but by the nonlinearity in the coefficients across categories of the dependent variable. It is possible, for instance, for a positive coefficient ultimately to

22. Block log-likelihood tests are $-2(LL_{partial} - LL_{full})$, which is distributed χ^2, where *partial* represents the model specified without the variable in question and *full* represents the fully specified model.

Case Study: Voting in Canada

Table 1: Comparison of CNES Sample Reported Votes to Actual, 1997

	Quebec			Rest of Canada		
		Percentage:			Percentage:	
	Population	Actual	Reported	Population	Actual	Reported
Abstain	1,974,538	35.0	11.7	7,817,709	45.6	11.5
Liberal	1,342,567	24.0	33.0	3,651,710	21.3	31.9
PC	811,410	14.0	18.7	1,635,295*	9.5	14.0
NDP	71,558	1.0	29.0	1,362,951	8.0	11.7
Reform	10,767	0.0	0.0	2,502,313*	14.6	27.0
Bloc	1,385,821	25.0	33.7	0	0.0	0.0
Other	37,772	1.0		173,710	1.0	
Votes	3,659,895			9,325,979		
VAP	5,634,433			17,143,688		

Parties: Liberal, Progressive Conservative, New Democratic, Reform, Bloc Quebecois.

Case Study: Voting in Canada

- Concentrate on Quebec votes ignoring small categories (NDP, Reform, and Other).
- Given a covariate matrix $\mathbf{X}$, the probability that respondent i chooses party j over abstention is:

$$p_{ij} = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{k=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_k)}$$

- The product of the estimated coefficients and the explanatory variables for the i^{th} individual is equal to the log of the odds of i selecting choice j (either Liberal, Progressive Conservative, or Bloc) divided by the odds of abstaining:

$$\log\left[\frac{p(y_{ij})}{p(y_{i1})}\right] = \mathbf{X}_i \boldsymbol{\beta}_j,$$

so positive values of β_k mean that increasing values of variable X_k push the log ratio towards selecting category j over the baseline category and negative values push the log ratio towards selecting the baseline category over category j.

Case Study: Voting in Canada

Table 2: Multinomial Logit Model of Vote Choice in Quebec, 1997

		Lib. vs. Abs.		PC vs. Abs.		Bloc Q. vs. Abs.	
		Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
(Intercept)		-3.4290	2.2072	-4.4160	2.2288	-2.5920	2.8967
Party Identification	Liberal	2.1663	1.1634	1.2821	1.2509	0.7587	1.5699
	ProgCons	13.5499	0.9867	14.2191	0.7387	13.5175	0.9519
	Bloc	-0.4093	1.2220	0.2529	0.7783	1.4918	0.5265
Parties Necessary?	middle	0.3399	0.6148	0.8945	0.5770	-0.3459	0.4979
	unnec	-1.7035	1.0272	-0.9432	0.9215	-2.1489	0.6713
Party Feeling	Liberal	0.7028	0.1736	-0.2836	0.1321	-0.0642	0.1186
Thermometer	ProgCons	-0.1095	0.1848	0.7120	0.1629	-0.1230	0.1306
	Bloc	-0.2769	0.1158	-0.2269	0.1136	0.2861	0.0968
Retrospective	Same	-0.6117	0.5851	-0.7164	0.5476	-0.4706	0.5035
Econ.Eval.	Worse	-1.0261	0.8110	-1.2789	0.7943	-0.6272	0.6237
Contact MP?	Yes	0.9169	0.8431	0.8538	0.7840	1.5085	0.6900
Left Right scale	Center	-0.1115	0.8971	-0.0236	0.7687	-1.5642	0.6899
	Right	0.8628	0.9011	0.2749	0.8096	-1.3709	0.6994
	DK	0.8224	0.8376	-1.0568	0.7517	-0.9012	0.6266
District Level Comp.		2.3685	1.9841	3.8490	1.9964	1.0196	1.7152
Political Info. Level	Low	1.2535	1.0102	0.8225	1.1107	-0.3135	0.8680
	Medium	0.9584	1.0374	1.5101	1.1332	0.7571	0.9070
	High	0.3752	1.0389	0.8955	1.1516	-0.7448	0.9503
Female		1.2007	0.5901	0.9267	0.5345	0.4088	0.4625
French		-2.1049	0.8435	-0.7949	0.8873	2.6080	2.3694

More On Tests for Violations of IIA

- Label $\hat{\boldsymbol{\beta}}_u$ the unrestricted model coefficient estimate, and $\hat{\boldsymbol{\beta}}_r$ the restricted model with at least one choice removed.

- All three test statistics are chi-square distributed under the null assumption of IIA with the degrees of freedom equal to the number of choices in the restriction set.

- The McFadden, Train, Tye (1981) test determines whether a null hypothesis of IIA should be rejected:

$$MTT = -2\left[L_r\left(\hat{\boldsymbol{\beta}}_u\right) - L_r\left(\hat{\boldsymbol{\beta}}_r\right)\right]$$

  where $L_r()$ denotes the likelihood function for the resticted model (the only one used).

- Recall that the log-likelihood function used here is:

$$\ell(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J|\mathbf{X}) = \sum_{j=1}^{J}\sum_{y_i=j}\exp(\mathbf{X}_i\boldsymbol{\beta}_j) - \sum_{l=1}^{n}\log\left(1 + \sum_{j=1}^{J}\exp(\mathbf{X}_\ell\boldsymbol{\beta}_j)\right),$$

- But this is known to be biased towards failing to reject the null hypothesis.

More On Tests for Violations of IIA

- Small and Hsiao (1985) avoid this bias by randomly splitting the sample into two roughly equal subsamples: A and B, estimate the *unrestricted* model on both parts to get $\hat{\boldsymbol{\beta}}_{uA}$ and $\hat{\boldsymbol{\beta}}_{uB}$, they then create a weighted average coefficient

$$\hat{\boldsymbol{\beta}}_{uAB} = \left(\frac{1}{\sqrt{2}}\right)\hat{\boldsymbol{\beta}}_{uA} + \left(1 - \frac{1}{\sqrt{2}}\right)\hat{\boldsymbol{\beta}}_{uB},$$

  then eliminate one choice from the subsample B to estimate $\hat{\boldsymbol{\beta}}_{rB}$, and produce the test statistic

$$SH = -2\left[L_r\left(\hat{\boldsymbol{\beta}}_{uAB}\right) - L_r\left(\hat{\boldsymbol{\beta}}_{rB}\right)\right].$$

- The Hausman and McFadden (1984) test determines whether a null hypothesis of IIA should be rejected:

$$HM = \left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_u\right)'\left[\hat{\text{Var}}(\hat{\boldsymbol{\beta}}_r) - \hat{\text{Var}}(\hat{\boldsymbol{\beta}}_u)\right]^{-1}\left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_u\right)$$

- Each of these are "weak" tests: the actual probability of rejecting the null hypothesis can be different than the nominal alpha level (Long and Freese 2006).

Case Study: Voting in Canada: HM Tests

- The model is reestimated several times, successively eliminating each choice alternative.
- The vectors of coefficients in the exclusionary models are then compared to the vectors of coefficients in the full model, with resulting Chi-Square tests.
- The df difference is one since we are excluding one party at time.
- Note that HM is a vector of length two since there are two parties left after the exclusion and making abstention the baseline (it's effect is produced by moving the baseline to Liberal).

Table 3: Hausman-McFadden Test of IIA Assumption in Quebec Model

	Liberal	PC	Bloc
Exclude Liberals	–	0.00	0.12
Exclude PC	0.00	–	0.00
Exclude Bloc	0.00	0.00	–
Exclude Abstention	baseline	0.00	0.00

- In every case we reject IIA, meaning that Quebec has a choice set effect even though Bloc Quebec dominates.

Nested Logit

- Start with a first level choice, $i = 1, 2, \ldots C$, followed by a subsequent second level choice, $j = 1, 2, \ldots, N_i$, which is nested in the first (note the subscript on N_i).

- McFadden (1978) uses the example of picking community to live in then picking a house to purchase.

- Write the standard latent utility function as $U_{ij} = \boldsymbol{\theta}_{ij} + \epsilon_{ij}$, and assume $\epsilon_{ij} \sim$ Weibull, then we can go on to define:
$$\boldsymbol{\theta}_{ij} = \mathbf{X}_{j|i}\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma}$$
where $\mathbf{X}_{j|i}$ is a set of covariates that apply at both levels of nesting and $\mathbf{Z}_i$ is a different set of covariates for the higher level only (as the subscripts imply).

- The unconditional choice probability is:
$$\begin{aligned} p(\text{house}_j, \text{neighborhood}_i) &= p(y_{ij}) \\ &= \frac{\exp(\mathbf{x}_{j|i}\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma})}{\sum_{i=1}^{C}\sum_{j=1}^{N_i}\exp(\mathbf{x}_{j|i}\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma})}. \end{aligned}$$
where the first sum is over communities and the second sum is over houses in each community.

Nested Logit

- This joint decision is not available directly by the assumptions of the model so we obtain it from the conditional probability of choosing alternative j given a choice of alternative i, and the unconditional probability of choosing alternative i in the first place:

$$p(y_{ij}) = p(y_{j|i})p(y_i) = \left(\frac{\exp(\mathbf{X}_{j|i}\boldsymbol{\beta})}{\sum_{k=1}^{N_i} \exp(\mathbf{X}_{k|i}\boldsymbol{\beta})} \frac{\exp(\mathbf{Z}_i\boldsymbol{\gamma} + I_j)}{\sum_{k=1}^{C} \exp(\mathbf{Z}_i\boldsymbol{\gamma} + I_j)} \right)$$

defining:

$$I_j = \log \sum_{k=1}^{N_i} \exp(\mathbf{X}_{k|i}\boldsymbol{\beta})$$

- Estimation Method 1: perform the estimate $p(y_{ij})$ by first estimating $\boldsymbol{\beta}$ from the lower level conditional logit $p(y_{j|i})$ and then estimate the full model plugging in these values (Maddala 1983).

- Estimation Methods 2: Full Information Maximum Likelihood,

$$\ell(\boldsymbol{\beta}, \mathbf{Z}) = \sum_{i=1}^{n} \log(p(\text{house}_j|\text{neighborhood}_i)p(\text{neighborhood}_i))$$

R Packages for Nested Logit

- mlogit
- VGAM
- RSGHB
- mnlogit
- See "Multinomial logit models in R." (Yves Croissant) http://www.r-project.org/conferences/useR-2009/abstracts/pdf/Croissant.pdf.
- And Keith Train's exercises https://cran.r-project.org/web/packages/mlogit/vignettes/Exercises.pdf.

Simple Example of Nested Logit

- Back to the dataset from Bill Greene's econmetric text (2008 p.730): four possible transport modes, the ground nest with bus, train and car modes, and the fly nest with plane.

- A data frame containing 840 observations on 4 modes for 210 individuals: `individual` factor indicating individual with levels 1 to 200, `mode` factor indicating travel mode (car, air, train, bus), `choice` factor indicating choice (no, yes), `wait` terminal waiting time (0 for car), `vcost` vehicle cost component, `travel` travel time in the vehicle, `gcost` generalized cost measure, `income` household income, `size` party size.

```
lapply(c("mlogit","AER"),library, character.only=TRUE)
data(TravelMode)
head(TravelMode)
  individual  mode choice wait vcost travel gcost income size
1          1   air     no   69    59    100    70     35    1
2          1 train     no   34    31    372    71     35    1
3          1   bus     no   35    25    417    70     35    1
4          1   car    yes    0    10    180    30     35    1
5          2   air     no   64    58     68    68     30    2
6          2 train     no   44    31    354    84     30    2
```

Simple Example of Nested Logit

- Shape a `data.frame` for using the `mlogit` function again:

```
TravelMode <- mlogit.data(TravelMode, choice = "choice", shape = "long",
    alt.var = "mode", chid.var = "individual")
```

  where: `shape = "long` if each row is an alternative or `shape = "wide"` if each row is an observation, `alt.var` is the name of the variable that contains the alternative index (the default name is `alt`), and `chid.var` the variable that contains the choice index.

- Since there is only one alternative in the air side of the nested part it must be coupled for the model to be identified:

```
TravelMode$avinc <- with(TravelMode, (mode == "air") * income)
```

- Now run the model:

```
nested.out <- mlogit(choice ~ wait + gcost + avinc, TravelMode, reflevel = "car",
     nests = list(fly = "air", ground = c("train", "bus", "car")),
     unscaled = TRUE)
```

Simple Example of Nested Logit

```
summary(nested.out)
Frequencies of alternatives:                bfgs method
    car     air   train     bus             17 iterations, 0h:0m:0s
0.28095 0.27619 0.30000 0.14286             g'(-H)^-1g = 1.02E-07
                                            gradient close to zero
Coefficients :
                  Estimate Std. Error t-value  Pr(>|t|)
air:(intercept)    6.042373   1.331325  4.5386 5.662e-06
train:(intercept)  5.064620   0.676010  7.4919 6.795e-14
bus:(intercept)    4.096325   0.628870  6.5138 7.328e-11
wait              -0.112618   0.011826 -9.5232 < 2.2e-16
gcost             -0.031588   0.007434 -4.2491 2.147e-05
avinc              0.026162   0.019842  1.3185   0.18732
iv.fly             0.586009   0.113056  5.1833 2.180e-07
iv.ground          0.388962   0.157904  2.4633   0.01377

Log-Likelihood: -193.66
McFadden R^2:  0.31753
Likelihood ratio test : chisq = 180.21 (p.value = < 2.22e-16)
```

Simple Example of Nested Logit

- Interpreting the `iv` coefficient estimates, which are called "log-sum coefficients."
- The log-sum cofficient in a regular multinomial is 1, so this is a test of the nesting.
- The t-test is given by:

```
( (coef(nested.out)['iv.fly']-1)/sqrt(vcov(nested.out)['iv.fly', 'iv.fly']) )
   iv.fly
-3.661813
```

  which is outside of $(-1.96 : 1.96)$ so we can reject the null hypothesis that the first nesting is not required.

- We can also do a likelihood ratio test since ML is nested in NL:

```
logit.out <- update(nested.out, nests = NULL)
lrtest(nested.out, logit.out)
Model 1: choice ~ wait + gcost + avinc
Model 2: choice ~ wait + gcost + avinc
  #Df  LogLik Df  Chisq Pr(>Chisq)
  1   8 -193.66
  2   6 -199.13 -2 10.944   0.004202
```

Applying Nested Logit to the Rest of Canada (non-Quebec)

- The "lower model" posits a first choice to "go right" with either of the two right-of-center parties: Reform and Progressive Conservative.

- Binomial choice is estimated with a logit model with the following explanatory variables: age, female, satisfaction with democracy, rural, respondents' beliefs about whether parties are necessary, retrospective economic evaluation, the effectiveness of the Reform and PC candidates in the district, and a dummy variable indicating whether an incumbent Reform MP is seeking reelection.

Applying Nested Logit to the Rest of Canada (non-Quebec)

Table 4: Lower Level Equation, ROC

(Reform and PC Voters only)			
		Coefficient	Std. Err.
(Intercept)		0.013531	0.571883
Age		-0.008672	0.006410
Female		-0.451533	0.209709
Satisfaction with Democracy	fairly	0.143564	0.357033
	not very	0.606388	0.405564
Rural		-0.066969	0.238153
Parties Necessary?	middle	0.569105	0.267299
	unnec	0.779037	0.385055
Economic Evaluation	neither	0.434635	0.235344
	bad	0.125053	0.299842
Effective Reform		1.223614	0.295573
Effective PC		-0.806035	0.238181
Incumbent Reform		0.758984	0.268281

Applying Nested Logit to the Rest of Canada

- The above coefficients are multiplied by the explanatory variables in the lower level model for Reform and PC voters to create an instrumental variable (which has a value of zero for Liberal and NDP voters and abstainers)
- This is then included in the upper level model of voter choice, along with feeling thermometers for four parties, self-placement on a left/right scale, union membership, electoral competition in the district, level of political information, education, and a dummy equal to one for either francophones or allophones.
- The results of the upper level multinomial logit model are estimated as a non-nested choice between abstention, Liberal, NDP, or a right party (either PC or Reform).

		Lib. vs. Abs.		NDP vs. Abs.		Right vs. Abs.	
		Coef.	Std.Err.	Coef.	Std.Err.	Coef.	Std.Err.
(Intercept)		-1.7811	0.7158	-2.5417	0.9144	-1.7483	0.7586
Feeling	Liberal	0.4663	0.0561	-0.2313	0.0635	-0.2528	0.0524
Thermometers	NDP	-0.1352	0.0484	0.4497	0.0594	-0.1638	0.0490
	ProgCons	-0.0203	0.0505	0.0008	0.0627	0.3204	0.0531
	Reform	-0.1560	0.0394	-0.1237	0.0518	0.1788	0.0380
Left/Right	Center	-0.3854	0.3476	-0.8034	0.3974	-0.4117	0.3756
	Right	0.0079	0.3678	-1.1196	0.4602	-0.1152	0.3866
	DK	-0.5421	0.3222	-1.4672	0.3727	-0.5713	0.3529
Union		-0.1088	0.2101	0.3620	0.2575	-0.1057	0.2141
District Comp.		0.8635	0.7415	1.3657	0.9389	1.9913	0.7826
Political Info.	Low	-0.2718	0.2938	0.8329	0.4552	0.5836	0.3036
	Medium	0.4919	0.3044	1.3362	0.4760	1.1394	0.3270
	High	1.1312	0.3630	2.1844	0.5168	2.2131	0.3826
Education	High School	0.0580	0.2945	0.3312	0.3910	0.0659	0.2925
	Post HS	-0.0434	0.2799	0.0846	0.3763	-0.1059	0.2782
	University	0.6834	0.3369	0.8197	0.4251	0.7113	0.3476
French or Allo		0.0017	0.3554	-0.6594	0.5341	-0.0964	0.3977
Instrument		0.1559	0.2206	-0.0278	0.2799	1.6063	0.1866

Nested Logit Model of Vote Choice for Rest of Canada, 1997

Protectionism Model

THE INTERNATIONAL TRADE COMMISSION AND THE POLITICS OF PROTECTIONISM

WENDY L. HANSEN
State University of New York
Stony Brook

I analyze the governmental regulation of internationally traded goods produced by U.S. industries. General theories of regulation—most notably "capture" theories and the theory of "congressional dominance"—are used to analyze the decision-making behavior of the U.S. International Trade Commission, which plays a major role in approving and providing tariffs, quotas, and various types of nontariff trade barriers sought by these industries. Unlike previous studies, this one simultaneously accounts for both the supply and demand sides of trade regulation. This work seeks to predict, on a basis of domestic politics, the factors that affect the demand for, and supply of, trade protection for U.S. industries. The methodology consists of applying a nested logit framework to capture the decision behavior of the International Trade Commission and industries simultaneously. The analysis shows that industries do appear to self-select themselves in applying for protection from the International Trade Commission. In light of these findings, it appears that trade protection is subject to domestic political forces similar to those affecting other regulatory policy areas.

Industries in the United States enjoy varying degrees of protection from foreign competition. While economic reasons may exist to justify some of these differences in protection, most economists and political scientists agree that one needs to look at the politics behind protective legislation to understand industry-specific differences in government assistance. My purpose here is to try to explain the varying levels of protection across industries by focusing on factors that affect both the supply of, and demand for, the regulation of trade. What circumstances lead industries to request protection and what factors affect the government's decision to supply that protection or not? Both industries and the government presumably have incentives to pursue utility-maximizing courses of action. On the demand side, when an industry seeks a higher tariff, the benefits from that tariff presumably outweigh the costs of applying and lobbying for protection. On the supply side, when the government chooses to protect an industry, the political benefits in terms of votes or contributions presumably exceed the loss of support from those harmed by the policy.

Given the incentives of the actors, I seek to predict, on the basis of domestic politics, the factors that explain industry and government decisions on trade matters. Why, for example, did the electric golf cart industry get higher tariffs in 1976 when the hand tool industry was turned down? In 1983, frozen orange juice makers got protection, but the canned mushroom industry was unsuccessful.

Protectionism Model

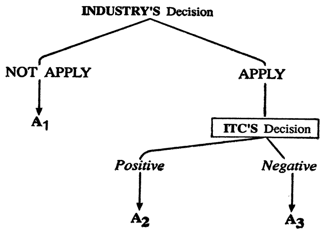
American Political Science Review Vol. 84

some form of relief through ITC action in a given year and zero to those that were denied relief or protection. About 40% of the industries that filed petitions in these years were granted some form of regulatory relief.

On the demand side, there are 425 four-digit SIC manufacturing industries included in this study with data for the same period, 1975–1984. These are all industries that compete on some level with foreign imports for a share in the U.S. market. A value of one is assigned in cases where at least one petition is filed for any industry in a given year and zero to cases where no petitions are filed. Industries filed a petition with the ITC in only about 7% (290) of the 4,250 entries across the 10 years. (Again, the number of observations was reduced in the estimations to only 2,903 cases because of missing data on various exogenous variables.) However, it should be noted that 133 of the 425 industries (31%) overall actually did file petitions at some time during this 10-year period, so the sample does represent a substantial number of industries.

All of the data used in this study are coded at the four-digit SIC level except the elasticities, where three-digit level data are used instead, with entries repeated for each corresponding four-digit code. Details on the independent variables used to explain the supply of, and demand for, the regulation of international trade are provided in Appendix A.

Figure 1. Two-Stage Decision Process



A nested logit model (McFadden 1978) is applied to the study of the demand for, and supply of, trade regulation to determine whether industries base their decisions to apply on their perception of the expected utility of getting protection. The demanders (industries) face the binary choice of whether or not to undergo the costs of applying and pressuring for trade regulation. The supplier, the ITC, makes the binary decision of whether or not to grant regulatory benefits to each of the applicants. This study aims to determine whether self-selection is a problem in predicting the probability of an industry getting protection; that is, Do industries self-select themselves in choosing whether or not to apply? By comparing the utility of not applying with the maximum expected utility that can be derived from filing an application, an industry can make a rational decision as to the usefulness of seeking protection from the ITC. By using a nested logit model, one can determine whether or not self-selection occurs.

Figure 1 illustrates the postulated structure of the model for the actors' choices. The model assumes that the regulator's decision is conditional on an industry's choice of applying. Stage 1 is the industry's decision of whether or not to file an application for protection. Stage 2 is the ITC's decision of whether or not to grant protection to the industry. The nested logit model was chosen because it characterizes the two-stage decision process well and allows for dependence among the attributes of the alternatives.

Suppose the utility of final outcome *ri* is represented by U_{ri}. The utility can be rewritten as the sum of the observable components V_{ri} and the unobservable disturbances ϵ_{ri}:

$$U_{ri} = V_{ri} + \epsilon_{ri} \text{ for } i, r = 1, 2,$$

Protectionism Model

Table 1. Coefficient Estimates for the Nested Logit Model

	Determinants of ITC Decisions[a]		Determinants of Industry Decisions[b]	
Variable	Coefficient	t-Statistic	Coefficient	t-Statistic
Constant	-3.68	-2.47*	-2.12	-9.47*
Elasticity of demand	-.31	-.96	—	—
Industry employment	1.14	.67**	—	—
Ways and Means Democrats	-.12	-.98	—	—
Ways and Means Republicans	.20	.64**	—	—
Trade subcommittee Democrats	.61	3.15*	—	—
Trade subcommittee Republicans	-.75	-1.93*	—	—
Ways and Means chair	1.28	2.74*	—	—
Ways and Means ranking member	.09	.14	—	—
Trade subcommittee chair	-.25	-.50	—	—
Trade subcommittee ranking member	-.11	-.19	—	—
Capacity utilization	.95	.62	—	—
U.S. trade deficit	1.38	2.46*	—	—
Industry concentration ratios	-.48	-.05	-3.81	-1.07
Percentage change in industry employment	-2.07	-1.12**	-1.27	-1.82*
Percentage change in market share	7.60	1.74*	-.90	-2.81*
Tariff rate	1.89	1.45	-1.19	-2.58*
Inclusive value	—	—	.18	3.29*
Number of cases	205		2,903	
Percentage correctly predicted	72		92.97	

[a]The dependent variable is the ITC decision: 1 = *protection*, 0 = *no protection*. There were 80 positive decisions and 125 negative decisions by the ITC.

[b]The dependent variable is the industry decision: 1 = *apply*, 0 = *not apply*. There were 205 industry applicants and 2,698 nonapplicants.

*$p \leq .05$, two-tailed test.

**Indicates $p \leq .05$ when the number of congressional representatives for each industry is replaced by a dummy variable.

sentative is a Democrat and a member of the trade subcommittee of Ways and Means, location of the industry in a district whose representative is the chair of Ways and Means, the U.S. trade deficit, size measured by employment, and the industry's percentage change in employment. Because employment is highly correlated with industry representation in the House of Representatives, size and percentage change in employment do not appear significant in Table 1. However, when the congressional representation variables (Ways and Means Democrats, trade subcommittee Democrats, Ways and Means Republicans, trade subcommittee Republicans) are replaced by dummies signifying representation by at least one committee member, size becomes significant at the 2.5% level, and percentage change in employment is significant at the 10% level. (Also, representation on the Ways and Means Committee by Republicans becomes significant at the 5% level when dummies are used here, but the significance of representation by Democrats disappears.) Presidential influence measured by party identification (not in the table) was insignificant with the t-statistic at −0.105.

These results indicate some degree of support for a pressure group model of regulation, especially congressional dominance. The size of the industry measured by the employment variable indicates that larger industries are more likely to get

Multinomial Probit

- The multinomial probit model uses the assumption of multivariate normal error terms.

- Adams (1997) showed that normally distributed errors emerge from very general assumptions in his simulation study. This is basically an expression of the persistence of the central limit theorem, but it highlights the fact that normally distributed errors are not only more tied to mathematical-statistics theory, they also emerge empirically.

- MNP is actually a much more restrictive model than MNL because it is non-identified without significant estimation restrictions.

- However, MNP does *not* require the IIA assumption (there are no log-odds of alternative probabilities excluding others).

Multinomial Probit

- Suppose there exist N respondents in the dataset with c choices observed for each respondent:

$$\boldsymbol{\omega}_i = [\omega_{i1}, \omega_{i2}, \ldots, \omega_{ic}],$$

  where all but one of these vector values is zero with the remaining value equal to one indicating individual selection.

- It is standard and convenient to assume that $\boldsymbol{\omega}_i$ is the observable manifestation of an underlying continuous measure of utility, $\mathbf{U}_i = [U_{i1}, U_{i2}, \ldots, U_{ic}]$, in which j^{th} value of $\boldsymbol{\omega}_i$ is equal to one because the associated latent measure has the greatest utility to person i of all alternatives: $U_{ij} > U_{ik}, \ \ \forall k \neq j$.

- So the ith person picks category j producing a value of 1 at this place in the vector $\boldsymbol{\omega}_i$ and zeros elsewhere because it has the largest value of U_{ij}.

Multinomial Probit

- We further assume that these utilities are generated by the distribution:

  $$\mathbf{U}_i \sim \mathcal{N}(\mathbf{Z}_i\boldsymbol{\gamma}, \Omega_{\mathbf{z}}),$$

  where: $\mathbf{Z}_i$ is a $c \times k$ data matrix, $\boldsymbol{\gamma}$ is a $k \times 1$ coefficient vector and $\Omega_{\mathbf{z}}$ is a $c \times c$ covariance matrix.

  That is, the underlying motivation for the model is multivariate Gaussian-normal:

  $$f(\mathbf{U}) = (2\pi)^{-n/2}|\Omega_{\mathbf{z}}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{U}-\mathbf{Z})'\Omega_{\mathbf{z}}^{-1}(\mathbf{U}-\mathbf{Z})\right]$$

- This is not identified in the same way as MNL, and it is again necessary to set a reference category choices comparatively (motivations in: Bunch 1991, Dansie 1985).

- Thus we reexpress from absolute utilities for person i, U_{ij}, to relative utilities, $y_{ij} = U_{ij} - U_{i1}$, where this relative to the arbitrary baseline category as in the MNL model.

Multinomial Probit

- This produces the assumed model for individual i:

$$\mathbf{y}_i \sim \mathcal{N}(\mathbf{X}_i\boldsymbol{\beta}, \Omega_{\mathbf{x}}), \qquad \text{where } \mathbf{X}_i \text{ is } (c-1)\times k, \ \ \boldsymbol{\beta} \text{ is } k \times 1, \ \ \Omega_{\mathbf{z}} \text{ is } (c-1)\times(c-1).$$

- The result of this specification is that the error terms in the model are now multivariate normal distributed, rather than Weibull distributed as in the MNL model.

- Now introduce a new variable $W_{ij} = I(y_{ij} > 0, y_{ij} = \max(y_{i\cdot}))$, and: $W_{i1} = 1$, $W_{i2:J} = 0$ if all values of y_{ij} are negative (McCulloch 1994). This indicator function makes the estimation of the coefficients much easier. The MNP likelihood is now the simple form:

$$\ell(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J) = \prod_{j=1}^{J}\prod_{i=1}^{N} \pi_{ij}^{W_{ij}},$$

  where π_{ij} is the probability that the i^{th} individual selects choice j with the obvious constraints that $\pi_{ij} > 0, \ \forall j$, and $\sum_{j=1}^{J} \pi_{ij} = 1$.

Multinomial Probit

- This model is still not identified because the scale of the relative utilities, y_{ij}, is indeterminate.
- Various authors in political science have dealt with this in various ways, some of which are quite restrictive:
  - Alvarez and Nagler (1995, 1998) and Lacy and Burden (1999) set all posterior variances to unity (Burden and Lacy also set one covariance equal to zero). The result of this change to unity along the diagonal is to make the covariance matrix a correlation matrix, which works well when the off-diagonal elements are of prime interest.
  - Quinn, Martin, and Whitford (1999 while WashU grad students) are less restrictive and merely confine the first diagonal term in the covariance matrix to be unity.

*When Politics and Models Collide: Estimating Models of Multiparty Elections**

R. Michael Alvarez, *California Institute of Technology*
Jonathan Nagler, *University of California, Riverside*

Theory: The spatial model of elections can better be represented by using conditional logit models which consider the position of the parties in issue spaces than by multinomial logit models which only consider the position of voters in the issue space. The spatial model, and random utility models in general, suffer from a failure to adequately consider the substitutability of parties sharing similar or identical issue positions.
Hypotheses: Multinomial logit is not necessarily better than successive applications of binomial logit. Conditional logit allows for considering more interesting political questions than does multinomial logit. The spatial model may not correspond to voter decision-making in multiple party settings. Multinomial probit allows for a relaxation of the IIA condition and this should improve estimates of the effect of adding or removing parties.
Methods: Comparisons of binomial logit, multinomial logit, conditional logit, and multinomial probit on simulated data and survey data from multiparty elections.
Results: Multinomial logit offers almost no benefits over binomial logit. Conditional logit is capable of examining movements by parties, whereas multinomial logit is not. Multinomial probit performs better than conditional logit when considering the effects of altering the set of choices available to voters. Estimation of multinomial probit with more than three choices is feasible.

1. The Theory and the Practice of Issue Voting Models

The spatial model of voting has been a dominant paradigm in the voting literature over the past 25 years (Davis, Hinich, and Ordeshook 1970; Downs 1957; Enelow and Hinich 1984), supplanting the "funnel of causality" (Campbell et al., 1960) which had a brief reign beginning around 1960.

*This is one of many joint papers by the authors on multiparty elections, the ordering of their names reflects alphabetic convention. Earlier versions of this research were presented at the Annual Meetings of the American Political Science Association, Chicago, IL, September 1995 and at the Annual Political Methodology Summer Conference, Indianapolis, July, 1995. We thank John Aldrich, Nathaniel Beck, Simon Jackman, John Jackson, Jonathan Katz, Gary King, Dean Lacy, Eric Lawrence, Jan Leighley, Will Moore, Mitch Sanders, and Guy Whitten for their comments on earlier versions of this research, and Methodology Conference participants for their input. We also thank participants of the Southern California Political Economy Group for their discussion of this research on November 17, 1995 at the University of California-Irvine, and participants in the Second CIC Interactive Video Methods Seminar which was broadcast from the University of Minnesota on October 25, 1996. Alvarez thanks the John M. Olin Foundation for support of his research. Nagler thanks the NSF for grant SBR-9413939. Comments may be directed to the authors at: DHSS 228-77, California Institute of Technology, Pasadena, Ca 91125, Internet: rma@crunch.caltech.edu; and Department of Political Science, University of California, Riverside, CA 92521-0118, Internet: nagler@wizard.ucr.edu, respectively.

American Journal of Political Science, Vol. 42, No. 1, January 1998, Pp. 55–96

long thought to be quite predictable, given that voting seemed to revolve primarily around social and religious cleavages in the electorate (Daalder 1966; Lijphart 1968). However, many scholars have begun to reexamine electoral politics in the Netherlands given the sudden rise in electoral volatility in recent decades (e.g., Middendorp and Tanke 1990; Van Der Eijk and Niemoller 1987; Whitten and Palmer 1996). Unfortunately few of the existing studies on electoral politics in the Netherlands have utilized models which do not assume that IIA holds for voters.[24]

Furthermore, there are a number of important questions which need to be answered about political change in the Netherlands. Most immediate is determining what has produced the dramatic increase in electoral volatility seen in the Netherlands since the mid-1960s (Bartolini and Mair 1990). Many scholars attribute this to the breakdown of "consociationalism" (Van Der Eijk and Niemoeller 1983). But what is fueling this breakdown? What factors are driving voter choice in contemporary Dutch politics? While some have argued that ideology is now determining voter choice (Van Der Eijk and Niemoller 1987), others have asserted that retrospective economic voting is the key to understanding recent elections in the Netherlands (Middendorp and Tanke 1990), and others have found the explanation somewhere in between (Whitten and Palmer 1996).

We use the 1994 Dutch Parliamentary Election Study for our analysis. We are able to develop a set of independent variables which would allow for close examination of the factors which determined voting in this election (ideological positioning of the parties, views on materialist and post-materialist issues, retrospective economic views, as well as religious and social status). The survey data were rich enough to allow us to explore voting for five of the parties which received the greatest vote shares in the 1994 election: Christian Democratic Appeal (CDA, 22.2%), Labor Party (PvdA, 24.0%), Liberal Party (VVD, 19.9%), Democrats' 66 (D66, 15.5%), and Green Left (GL, 3.5%).[25]

[24]Quinn, Martin, and Whitford (1996) and Schofield et al. (1997) provide extensive analyses of the 1979 Dutch election using a different estimation technique than we utilize; their work employs the Gibbs sampling for estimating multinomial probit models (Albert and Chib 1993; McCulloch and Rossi 1994).

[25]The variables we used in our model of the 1994 Dutch election were taken from the *Dutch Parliamentary Election Study (DPES) 1994,* overseen by H. Ankers and E. V. Oppenhuis; this data is available from the ICPSR. The ideology variable we employ is coded as the absolute distance between the respondent and the mean ideological position of each party, with the latter estimated from the survey sample. We use variables measuring materialist and post-materialist values; each of these variables are factor scales, where positive values indicate strong materialist or post-materialist values, constructed from a two dimensional principal components analysis of responses to 17 questions included in the DPES (variables v497–v513). To measure economic perceptions, we use three variables, each of which is coded so that the high category expresses favorable responses about the

Table 9. Simulated Multinomial Probit Estimates, 1994 Dutch Election

Independent Variables		PVDA/GL	CDA/GL	VVD/GL	D66/GL
Ideology	–.37*				
	(.06)				
Constant		.67	–1.1	1.0	2.6**
		(1.2)	(.76)	(1.3)	(1.3)
Materialism		–2.4*	–2.2*	–2.5*	–1.5
		(.96)	(1.1)	(1.0)	(.98)
Postmaterialism		.58*	1.1*	1.1*	.54*
		(.19)	(.20)	(.19)	(.18)
Economy		.51*	.72*	.28	.16
		(.22)	(.23)	(.25)	(.23)
Employment		.46*	.34*	.21	.27**
		(.16)	(.16)	(.16)	(.15)
Personal Finances		–.10	–.13	–.40**	–.52**
		(.20)	(.22)	(.21)	(.20)
Catholic		–.31	.86*	–.07	–.24
		(.29)	(.32)	(.86)	(.28)
Reform		.23	–1.1	.30	.02
		(.23)	(.76)	(.24)	(.28)
Calvinist		–.81	1.6**	–.07	–.05
		(.88)	(.82)	(.86)	(.81)
Age		1.50*	1.8*	1.0*	–.30
		(.41)	(.42)	(.40)	(.44)
Education		–.15	–.05	–.03	–.18
		(.17)	(.17)	(.17)	(.17)
Gender		–.40*	–.55*	–.41*	–.29
		(.18)	(.20)	(.18)	(.19)
Income		.40	.86*	1.1*	.66*
		(.26)	(.28)	(.28)	(.26)
Urban		.10	.16	.03	.08
		(.10)	(.10)	(.11)	(.10)
Manual workers		–.21	–.73*	–.38	–.29
		(.29)	(.33)	(.29)	(.31)
Union members		.25	–.01	–.19	–.00
		(.27)	(.30)	(.29)	(.28)
$\delta_{CDA,VVD}$	.47				
	(.31)				
$\delta_{PVDA,CDA}$	.54*				
	(.18)				
$\delta_{PVDA,VVD}$	.29				
	(.26)				
$\delta_{PVDA,D66}$	.007				
	(.29)				
Number of Obs	901				
LL	–931.0				

Standard Errors in parentheses. * indicates significance at 95% level; ** indicates significance at 90% level.

R Packages for Multinomial Probit

- `MNP`, Imai and van Dyke
- `endogMNP`, Burgette (an extension of Imai and van Dyke)
- `bayesem`, Rossi
- `MCMCpack`, Martin, Quinn, Park
- `mlogit`, Croissant

Transportation Example of Multinomial Probit

- Use another transportation example from the `mlogit` package with 453 commuters choosing bus, car, carpool, or rail.
- Utility differences are computed respective to the reference level of the response (default=bus).
- The 3×3 covariance matrix is now estimated to get:

$$\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}', \qquad \boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 \\ \sigma^2_{32} & \sigma^2_{33} & 0 \\ \sigma^2_{42} & \sigma^2_{43} & \sigma^2_{44} \end{bmatrix}$$

where:

CoVar(car.carpool)	σ^2_{32}
CoVar(car.rail)	σ^2_{42}
CoVar(carpool.carpool)	σ^2_{33}
CoVar(carpool.rail)	σ^2_{43}
CoVar(rail.rail)	σ^2_{44}

- Since the first element of this matrix is set to $\sigma^2_{22} = 1$ then the model is identified.

Transportation Example of Multinomial Probit

- Get and condition the data:

```
library(mlogit); data("Mode")
head(Mode)
```

```
  choice cost.car cost.carpool cost.bus cost.rail  time.car time.carpool time.bus time.rail
1    car    1.507       2.3356    1.801     2.359   18.503       26.338    20.87     30.03
2   rail    6.057       2.8969    2.237     1.855   31.311       34.257    67.18     60.29
3    car    5.795       2.1375    2.576     2.747   22.547       23.255    63.31     49.17
4    car    1.869       2.5724    1.904     2.268   26.090       29.896    19.75     13.47
5    car    2.499       1.7220    2.686     2.974    4.699       12.414    43.09     39.74
6    car    4.727       0.6242    1.848     2.310    3.073        9.223    12.83     43.54
```

```
TravelMode2 <- dfidx(Mode, choice="choice", varying=2:9)
```

where `varying` indexes the variables that are alternative specific.

- Run the model

```
MNP2 <- mlogit(choice~cost+time, TravelMode2, seed = 20, R = 100, probit = TRUE)
```

where `R` is the number of function evaluation for the gaussian quadrature method, you need to set a random seed since the estimation is done by simulation.

Transportation Example of Multinomial Probit

```
summary(MNP2)
Frequencies of alternatives:          bfgs method          20 iterations, 0h:0m:35s
   bus     car carpool    rail        g'(-H)^-1g = 7.71E-07  gradient close to zero
0.1788  0.4812  0.0706  0.2693
Coefficients:
                    Estimate Std. Error t-value Pr(>|t|)
car:(intercept)      1.83087    0.25064    7.30  2.8e-13
carpool:(intercept) -1.28168    0.56778   -2.26  0.02399
rail:(intercept)     0.30935    0.11517    2.69  0.00723
cost                -0.41344    0.07316   -5.65  1.6e-08
time                -0.04666    0.00683   -6.83  8.2e-12
car.carpool          0.25997    0.38503    0.68  0.49955
car.rail             0.73649    0.21457    3.43  0.00060
carpool.carpool      1.30789    0.39167    3.34  0.00084
carpool.rail        -0.79818    0.34637   -2.30  0.02120
rail.rail            0.43013    0.48746    0.88  0.37757

Log-Likelihood: -348      McFadden R^2:  0.36
Likelihood ratio test : chisq = 392 (p.value = <2e-16)
```

Transportation Example of Multinomial Probit

- The non-intercept coefficients are not easy to interprete from the `mlogit` output as given:

```
cost                    -0.41344     0.07316    -5.65   1.6e-08
time                    -0.04666     0.00683    -6.83   8.2e-12
```

- We need to do some comparisons to put them in context, so $\hat{\beta}_{\text{time}}/\hat{\beta}_{\text{cost}}$ is $-0.04666/-0.41344 = 0.1129$, which means that we get roughly one-tenth of a Euro value for a minute of traveling, equivalently one Euro value for roughly 9 minutes of traveling.

- These are the estimated covariance terms from the table:

```
                    Estimate Std. Error t-value Pr(>|t|)
car.carpool          0.25997     0.38503    0.68  0.49955
car.rail             0.73649     0.21457    3.43  0.00060
carpool.carpool      1.30789     0.39167    3.34  0.00084
carpool.rail        -0.79818     0.34637   -2.30  0.02120
rail.rail            0.43013     0.48746    0.88  0.37757
```

Transportation Example of Multinomial Probit

- Look at the covariance matrix:

```
L <- matrix(0,ncol=3,nrow=3);  L[!upper.tri(L)] <- c(1, coef(MNP2)[6:10])
L %*% t(L)
      [,1]   [,2]   [,3]
[1,] 1.000  0.260  0.736
[2,] 0.260  1.778 -0.852
[3,] 0.736 -0.852  1.365
```

where the order is: car, carpool, rail.

- Compare predictions to observed proportions:

```
predict(MNP2)
    bus     car carpool    rail
0.14777 0.57268 0.08057 0.25607

MNP2$freq/sum(MNP2$freq)
    bus     car carpool    rail
0.17881 0.48124 0.07064 0.26932
```

Imai and van Dyke's Package

```
library(MNP)
# LOAD THE JAPANESE ELECTION DATA N THE PACKAGE
data(japan)
res2 <- mnp(cbind(LDP, NFP, SKG, JCP) ~ gender + education + age, data = japan,
        verbose = TRUE, n.draws=500000, burnin=250000)


The base category is ‘JCP’.

The total number of alternatives is 4.

The trace restriction is used instead of the diagonal restriction.

The dimension of beta is 12.

The number of observations is 418.

Improper prior will be used for beta.
```

```
Starting Gibbs sampler...
 10 percent done.
 20 percent done.
 30 percent done.
 40 percent done.
 50 percent done.
 60 percent done.
 70 percent done.
 80 percent done.
 90 percent done.
100 percent done.

# SUMMARIZE THE RESULTS
summary(res2)

Coefficients:
                     mean  std.dev.      2.5% 97.5%
(Intercept):LDP  0.801221  0.409471 -0.003176  1.60
(Intercept):NFP  1.107319  0.455469  0.213959  2.00
(Intercept):SKG  0.395675  0.363710 -0.315562  1.11
gendermale:LDP  -0.106309  0.150890 -0.403867  0.19
```

```
gendermale:NFP  -0.218796  0.165483 -0.544497  0.10
gendermale:SKG  -0.137354  0.134068 -0.402901  0.12
education:LDP   -0.103670  0.074942 -0.250094  0.04
education:NFP   -0.104305  0.082917 -0.266168  0.06
education:SKG   -0.002966  0.066188 -0.132904  0.13
age:LDP          0.013371  0.006007  0.001655  0.03
age:NFP          0.006847  0.006637 -0.006164  0.02
age:SKG          0.009560  0.005330 -0.000779  0.02

Covariances:
          mean std.dev.   2.5% 97.5%
LDP:LDP 0.9641   0.0511 0.8635  1.06
LDP:NFP 1.0161   0.0407 0.9307  1.09
LDP:SKG 0.6851   0.0564 0.5683  0.79
NFP:NFP 1.3637   0.0764 1.2204  1.52
NFP:SKG 0.7246   0.0618 0.5950  0.84
SKG:SKG 0.6721   0.0661 0.5418  0.80

Base category: JCP
Number of alternatives: 4
Number of observations: 418
```

```
Number of estimated parameters: 17
Number of stored MCMC draws: 250000

# CALCULATE THE PREDICTED PROBABILITIES FOR THE 10TH OBSERVATION
# AVERAGING OVER 100 ADDITIONAL MONTE CARLO DRAWS GIVEN EACH MCMC DRAW
pre2 <- predict(res2, newdata = japan[10,], type = "prob", n.draws = 100,
        verbose = TRUE)
There is one observation to predict. Please wait...
 10 percent done.
 20 percent done.
 30 percent done.
 40 percent done.
 50 percent done.
 60 percent done.
 70 percent done.
 80 percent done.
 90 percent done.
100 percent done.
apply(pre2$p[1,,],1,mean)
  JCP   LDP   NFP   SKG
0.108 0.356 0.328 0.208
```