Lecture Notes on Discrete Choice Models

Jonathan Nagler

1 Topics

1. Review the Latent Varible Setup For Binary Choice
   - Logit
   - Likelihood for Logit
   - Probabilities
2. **RUM:** The Random Utility Model formulation of choice.
   - Flexible
   - A strong (or weak) model of behavior.
3. Multinomial Logit
   - Likelihood
   - Probabilities
4. Conditional Logit
   - a More General Systemic Component of Model
   - Probabilities are Calculated the same as MNL
5. Measuring The Effects of Changes in X:
   - Analagous to OLS

- See: King - *Unifying Political Methodology*
- See: King/Tomz/Wittenberg (1998, *APSA Meeting*).
- See: Alvarez and Nagler 1995 ("Perot - 1992"; Table 4), first differences

6. Measuring the Effects of Changes in the Choices
   - See: Alvarez and Nagler 1998 ("Collide"; Table 5 - moving Labour Party).

7. Goodness of Fit
   - Percent correctly predicted in 2-choice case.
   - Baseline Prediction in 2-choice case.
   - Percent Correctly predicted in J-choice case:
     - Baseline Prediction
     - Classification Schemes

8. Independence of Irrelevant Alternatives (IIA)
   - What is IIA
   - IIA does not aggregate
   - McFadden test for IIA

9. Multinomial Probit (MNP)
   - Does not impose IIA.
   - Need to put restrictions on Σ.
   - Estimation via: Gauss, Gaussx, Limdep
   - Beyond 5 choices?

10. Some Equivalencies of Logit Models
    - MNL can be used to recover reduced form CL estimates.
    - MNL is equivalent to Binary Logit (under IIA).
11. Scobit
    - What if basic assumption of the shape of the response curve is wrong?
12. Heteroscedastic Probit
13. Selection Bias
    - Heckman
    - Dubin and Rivers

2 Latent Variable Setup: Binary Probit/Logit

$$
\begin{aligned}
y^* &= \beta' x + \epsilon && (1)\\
y &= 1 \; if \; y^* > 0 && (2)\\
y &= 0 \; if \; y^* \leq 0 && (3)
\end{aligned}
$$

Now assume F is the cumulative distribution for ϵ.

$$
\begin{aligned}
Prob(y = 1) &= Prob(y^* > 0) && (4)\\
&= Prob(\beta' x + \epsilon > 0)\\
&= Prob(\epsilon > -\beta' x)\\
&= 1 - Prob(\epsilon < -\beta' x)\\
&= 1 - F(-\beta' x) && (5)
\end{aligned}
$$

If F is symmetric about 0,

$$
\begin{aligned}
Prob(y = 1) &= 1 - F(-\beta' x)\\
&= F(\beta' x) && (6)
\end{aligned}
$$

If F is the logistic distribution this gives us logit, if F is the cumulative normal distribution we have probit.

In either case, logit or probit would recover consistent estimates of the parameter β.

Logistic distribution looks like:

$$
F(x) = \frac{1}{1 + e^{-x}} \qquad (7)
$$

So, if F is logistic:

$$\begin{aligned} Prob(y \ = \ 1) \ &= \ F(\beta' x) \\ &= \ \frac{1}{1 \ + \ e^{-\beta' x}} \\ &= \ \frac{e^{\beta' x}}{1 \ + \ e^{\beta' x}} \end{aligned}$$

So, if we could estimate β, then we could compute the quantity of interest (P).

We use maximum likelihood to compute β: we want to find β to maximize:

$$Pr(Y \mid \beta, \ X)$$

In the simple case, we have two possibilities: $y_i \ = \ 1$ or $y_i \ = \ 0$.

$$\begin{aligned} Pr(y_1, y_2, ..., y_N \mid \beta, X) \ = \ &\prod_{y_i=1} Pr(Y_i = 1 \mid \beta, X) \ * \\ &\prod_{y_i=0} Pr(Y_i = 0 \mid \beta, X) \end{aligned}$$

We work with something similar to the above: the likelihood function (L).

The following assumes F is symettric, and substitutes F for $Pr(Y_i = 1)$:

$$L = \prod_{y_i=1} F(X_i\beta \mid \beta, X) \prod_{y_i=0} (1 - F(X_i\beta \mid \beta, X))$$

$$= \prod F(X_i\beta \mid \beta)^{y_i} \prod (1 - F(X_i\beta \mid \beta))^{1-y_i}$$

We always work with $\log(L)$ - or the log-likelihood function.

$$LL = \sum y_i \ \ln(F(X_i\beta)) + \sum (1 - y_i)\ln((1 - F(X_i\beta)))$$

So, we take the first derivatives of the above expression with respect to β, set them equal to 0, and solve for $\hat{\beta}$.

Generally speaking, we do not solve for $\hat{\beta}$ analytically, we do it numerically. The log likelihood function is assumed to be well behaved, and we search for its maximum. **We can make mistakes.** We might find a local maximum that is not a global maximum. Or the likelihood function might be very flat, making our result sensitive to convergence criteria or the algorithim we choose.

3 Random Utility Models (RUM)

Assume the i^{th} individual's utility of the j^{th} choice is given as:

$$U_{ij} = V_{ij} \ + \ \epsilon_{ij} \tag{8}$$

where V_{ij} is a **systemic component of utility** and ϵ_{ij} is a **stochastic component of utility**.

Assume that the i^{th} individual chooses choice j iff:

$$U_{ij} \ > \ U_{ik} \quad \forall \ k \ \neq \ j \tag{9}$$

Notice that ϵ_{ij} is subscripted by i **and** j. We have one disturbance per respondent **per choice**.

Simple setup of V_{ij}:

$$V_{ij} \ = \ \beta_j X_{ij} \tag{10}$$

$$\tag{11}$$

This model is very flexible: it allows for more than 2 choices.

The model is a weak or strong **model of behavior.**

4 RUM Example 1: Multinomial Logit

$$V_{ij} = \psi_j \mathbf{X_i} \quad (12)$$
$$V_{ij} = \psi_{j0} + \psi_{j1}\mathbf{pid_i} + \psi_{j2}\mathbf{educ_i} + \psi_{j3}\mathbf{ideology_i} \quad (13)$$

$$\begin{aligned} V_{i1} &= \psi_{10} + \psi_{11}\mathbf{pid_i} + \psi_{12}\mathbf{educ_i} + \psi_{13}\mathbf{ideology_i} \\ V_{i2} &= \psi_{20} + \psi_{21}\mathbf{pid_i} + \psi_{22}\mathbf{educ_i} + \psi_{23}\mathbf{ideology_i} \\ V_{i3} &= \psi_{30} + \psi_{31}\mathbf{pid_i} + \psi_{32}\mathbf{educ_i} + \psi_{33}\mathbf{ideology_i} \end{aligned}$$

So:

$$\begin{aligned} U_{i1} &= \psi_{10} + \psi_{11}\mathbf{pid_i} + \psi_{12}\mathbf{educ_i} + \psi_{13}\mathbf{ideology_i} + \epsilon_{i1} \\ U_{i2} &= \psi_{20} + \psi_{21}\mathbf{pid_i} + \psi_{22}\mathbf{educ_i} + \psi_{23}\mathbf{ideology_i} + \epsilon_{i2} \\ U_{i3} &= \psi_{30} + \psi_{31}\mathbf{pid_i} + \psi_{32}\mathbf{educ_i} + \psi_{33}\mathbf{ideology_i} + \epsilon_{i3} \end{aligned}$$

ϵ are *iid*, Type I Extreme Value.

$$\begin{aligned} P_{i1} = & \quad Pr[(U_{i1} > U_{i2}) \quad \& \quad (U_{i1} > U_{i3})] \\ = & \quad Pr[(V_{i1} + \varepsilon_{i1} > V_{i2} + \varepsilon_{i2}) \quad \& \quad (V_{i1} + \varepsilon_{i1} > V_{i3} + \varepsilon_{i3})] \\ = & \quad Pr[(\varepsilon_{i2} - \varepsilon_{i1} \; < \; V_{i1} - V_{i2}) \quad \& \quad (\varepsilon_{i3} - \varepsilon_{i1} \; < \; V_{i1} - V_{i3})] \end{aligned}$$

Notice: ψ is indexed by j.

Normalization of One Set of ψ's

$$\begin{aligned} U_{i1} &= \psi_1 A_i + \varepsilon_{i1} \\ U_{i2} &= \psi_2 A_i + \varepsilon_{i2} \\ U_{i3} &= \psi_3 A_i + \varepsilon_{i3} \end{aligned}$$

$$\begin{aligned} P_{i1} &= Pr[(U_{i1} > U_{i2}) \quad \& \qquad (U_{i1} > U_{i3})] \\ &= Pr[(\psi_1 A_i + \varepsilon_{i1} > \psi_2 A_i + \varepsilon_{i2}) \quad \& \\ &\qquad\qquad (\psi_1 A_i + \varepsilon_{i1} > \psi_3 A_i + \varepsilon_{i3})] \\ &= Pr[(\varepsilon_{i2} - \varepsilon_{i1} < (\psi_1 - \psi_2) A_i) \;\; \& \\ &\qquad\qquad (\varepsilon_{i3} - \varepsilon_{i1} < (\psi_1 - \psi_3) A_i)] \end{aligned}$$

$$\begin{aligned} P_{i2} &= Pr[(\varepsilon_{i1} - \varepsilon_{i2} < (\psi_2 - \psi_1) A_i) \;\; \& \\ &\qquad\qquad (\varepsilon_{i3} - \varepsilon_{i2} < (\psi_2 - \psi_3) A_i)] \end{aligned}$$

$$\begin{aligned} P_{i3} &= Pr[(\varepsilon_{i1} - \varepsilon_{i3} < (\psi_3 - \psi_1) A_i) \;\; \& \\ &\qquad\qquad (\varepsilon_{i2} - \varepsilon_{i3} < (\psi_3 - \psi_2) A_i)] \end{aligned}$$

3 Quantities:

$$\begin{aligned} \Psi_1 - \Psi_2 &= X \\ \Psi_1 - \Psi_3 &= Y \\ \Psi 2 - \Psi_3 &= Z \end{aligned}$$

But:

$$Z = Y - X$$

Example:

$$\begin{aligned}\Psi_1 &= 7\\ \Psi_2 &= 4\\ \Psi_3 &= 0\end{aligned}$$

Yields:

$$\begin{aligned}\Psi_1 - \Psi_2 &= 3\\ \Psi_1 - \Psi_3 &= 7\\ \Psi 2 - \Psi_3 &= 4\end{aligned}$$

Same Result if:

$$\begin{aligned}\Psi_1 &= 24\\ \Psi_2 &= 21\\ \Psi_3 &= 17\end{aligned}$$

Yields:

$$\begin{aligned}\Psi_1 - \Psi_2 &= 3\\ \Psi_1 - \Psi_3 &= 7\\ \Psi 2 - \Psi_3 &= 4\end{aligned}$$

4.1 Probabilities:

We do not prove the following here; but it is true.

$$P_{ij} = \frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}} \tag{14}$$

Simple 2-Choice Case:

$$\begin{aligned} Pr(Y_i = 1) &= \frac{e^{\beta_1' X_i}}{e^{\beta_1' X_i} + e^{\beta_2' X_i}} \\ &= \frac{e^{\beta_1' X_i}}{e^{\beta_1' X_i} + 1} \end{aligned} \tag{15}$$

This looks like binary logit:

$$\begin{aligned} F(x) &= \frac{1}{1 + e^{-x}} \\ &= \frac{1}{1 + \frac{1}{e^x}} \\ &= \frac{1}{\frac{e^x + 1}{e^x}} \\ &= \frac{e^x}{e^x + 1} \end{aligned} \tag{16}$$

Table 2: Multinomial Logit and Binomial Logit Estimates
British Election - 1987 *(Alvarez and Nagler 1998)*

	Conservative/Alliance		**Labour/Alliance**	
	MNL	**BL**	**MNL**	**BL**
Intercept	-4.33*	-4.40*	4.55*	5.26*
	(.74)	(.76)	(.81)	(.86)
Defense	.14*	.17*	-.17*	-.19*
	(.03)	(.03)	(.03)	(.03)
Phillips Curve	.08*	.10*	-.03	-.05
	(.02)	(.03)	(.03)	(.03)
Taxation	.13*	.14*	-.06**	-.08*
	(.03)	(.03)	(.03)	(.04)
National.	.16*	.16*	-.16*	-.20*
	(.03)	(.03)	(.03)	(.03)
Redist.	.07*	.06*	-.08*	-.09*
	(.02)	(.02)	(.03)	(.03)
Crime	.08*	.08*	.02	.02
	(.03)	(.03)	(.02)	(.02)
Welfare	.11*	.12*	-.11*	-.10*
	(.02)	(.02)	(.03)	(.03)
South	-.12	-.06	-.41*	-.45*
	(.16)	(.17)	(.21)	(.22)
Midlands	-.26	-.26	-.12	-.15
	(.17)	(.17)	(.21)	(.21)
North	-.03	.03	.66*	.61*
	(.17)	(.18)	(.19)	(.20)
Wales	-.40	-.41	1.41*	1.46*
	(.35)	(.36)	(.31)	(.33)
Scot	-.36	-.42**	.68*	.61*
	(.25)	(.26)	(.25)	(.26)
Union	-.50	-.49*	.37*	.35*
	(.16)	(.16)	(.16)	(.17)
Public Employee	.04	.03	-.05	.03
	(.15	(.15	(.16	(.16
Blue Collar	.09	.14	.70*	.80*
	(.15)	(.16)	(.17)	(.17)
Gender	.29*	.33*	.04	-.03
	(.14)	(.14)	(.15)	(.16)
Age	.03	.03	-.21*	-.24*
	(.05)	(.05)	(.05)	(.05)
Homeowner	.31**	.26	-.55*	-52*
	(.18)	(.18)	(.17)	(.17)
Income	.07*	.07*	-.05	-.07*
	(.03)	(.03)	(.03)	(.03)
Education	-.81*	-.92*	-.54	-.65**
	(.31)	(.31)	(.35)	(.36)
Inflation	.28*	.31*	-.00	.05
	(.10)	(.11)	(.12)	(.12)
Taxes	.02	-.04	-.11	-.15*
	(.06)	(.07)	(.07)	(.07)
Unempl.	.30*	.30	.04	.08
	(.06)	(.06)	(.07)	(.08)
Number of Observations	2131	1494	2131	1172
Log Likelihood	-1500.8	-734.71	-1500.8	-588.27

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

Consider just a 2-choice comparison:

$$U_{i1} = \beta_1' X_i + \epsilon_{i1} \tag{17}$$
$$U_{i2} = \beta_2' X_i + \epsilon_{i2} \tag{18}$$

$$Pr(U_{i2} > U_{i1}) = Pr(\beta_1' X_i + \epsilon_{i1} < \beta_2' X_i + \epsilon_{i2})$$
$$= Pr(\epsilon_{i1} - \epsilon_{i2} < (\beta_2' - \beta_1') X_i) \tag{19}$$

Back to latent variable model:

$$Pr(Y_i = 1) = Pr(u_i < \beta' X_i) \tag{20}$$

So:

$$u_i \approx \epsilon_{i1} - \epsilon_{i2}$$
$$\beta \approx \beta_2 - \beta_1 \tag{21}$$

If we assume ϵ_{ij} are independent, identically distributed with Type 1 Extreme Value distribution, then $\epsilon_{i1} - \epsilon_{i2}$ is logistically distributed.

$$F(\epsilon_{ij}) = exp(e^{-\epsilon_{ij}}) \tag{22}$$

Assume $\beta_2 = 0$. This is just a normalization, we could assume $\beta_2 = 17$. The only thing that matters is $U_{i1} - U_{i2}$.

5 Conditional Logit

$$U_{ij} \;=\; \psi_j' A_i \;+\; \beta' X_{ij} \;+\; \epsilon_{ij} \tag{23}$$

where:

U_{ij} = utility of the i^{th} respondent for the j^{th} alternative.

A_i = characteristics of the i^{th} respondent.

X_{ij} = characteristics of the j^{th} alternative relative to the i^{th} respondent.

ψ_j = a vector of parameters relating the characteristics of a respondent to the respondent's utility for the j^{th} alternative.

β = a vector of parameters relating the relationship between the respondent and the alternative (X_{ij}) to the respondent's utility for the alternative.

ϵ_{ij} = random disturbance for the i^{th} respondent for the j^{th} alternative; *iid*, Type I Extreme Value.

Notice: ψ_j varies across choices.

Both conditional logit and multinomial logit models assume that the disturbances, ϵ_{ij}, are independent across alternatives.

6 RUM Example 2: Conditional Logit

$$
\begin{aligned}
V_{ij} &= \beta_1\, \mathbf{X_{ij}} + \psi_j \mathbf{A_i} \\
V_{ij} &= \beta_1\, \mathbf{issuedist_{ij}} + \psi_{j0} + \psi_{j1}\mathbf{pid_i} + \psi_{j2}\mathbf{educ_i}
\end{aligned}
$$

$$
\begin{aligned}
V_{i1} &= \beta_1\, \mathbf{issuedist_{i1}} + \psi_{10} + \psi_{11}\mathbf{pid_i} + \psi_{12}\mathbf{educ_i} \\
V_{i2} &= \beta_1\, \mathbf{issuedist_{i2}} + \psi_{20} + \psi_{21}\mathbf{pid_i} + \psi_{22}\mathbf{educ_i} \\
V_{i3} &= \beta_1\, \mathbf{issuedist_{i3}} + \psi_{30} + \psi_{31}\mathbf{pid_i} + \psi_{32}\mathbf{educ_i}
\end{aligned}
$$

So:

$$
\begin{aligned}
U_{i1} &= \beta_1\, \mathbf{issuedist_{i1}} + \psi_{10} + \psi_{11}\mathbf{pid_i} + \psi_{12}\mathbf{educ_i} + \epsilon_{i1} \\
U_{i2} &= \beta_1\, \mathbf{issuedist_{i2}} + \psi_{20} + \psi_{21}\mathbf{pid_i} + \psi_{22}\mathbf{educ_i} + \epsilon_{i2} \\
U_{i3} &= \beta_1\, \mathbf{issuedist_{i3}} + \psi_{30} + \psi_{31}\mathbf{pid_i} + \psi_{32}\mathbf{educ_i} + \epsilon_{i3}
\end{aligned}
$$

Table 4
Conditional Logit Estimates
British Election - 1987

	Conservative/Alliance		Labour/Alliance
Defense[a]		-.18*	
		(.02)	
Phillips Curve		-.11*	
		(.02)	
Taxation		-.16*	
		(.02)	
National.		-.18*	
		(.02)	
Redist.		-.08*	
		(.02)	
Crime		-.10*	
		(.05)	
Welfare		-.14*	
		(.02)	
Intercept	.82		2.53*
	(.69)		(.75)
South	-.15		-.44*
	(.17)		(.21)
Midlands	-.29**		.19
	(.17)		(.20)
North	-.06		.64*
	(.18)		(.19
Wales	.48		1.3*
	(.36)		(.31)
Scot	-.41		.69*
	(.25)		(.25)
Union	-.50*		.37*
	(.16)		(.16)
Public Employee	.09		-.02
	(.15)		(.16)
Blue Collar	.11		.70*
	(.15)		(.16)
Gender	.28*		.00
	(.14)		(.15)
Age	.02		-.22*
	(.05)		(.05)
Homeowner	.37*		-.54*
	(.18)		(.16)
Income	.07*		-.06
	(.03)		(.03)
Education	-.82*		-.61**
	(.32)		(.35)
Inflation	.28*		-.03
	(.10)		(.11)
Taxes	.01		-.10
	(.07)		(.07)
Unempl.	.28*		(.01
	(.06)		(.07)
N		2131	
Log Likelihood		-1477.6	

[a]The seven issues represent distance – absolute value – from the respondent to the mean of the party position.
Standard Errors in parenthesis. * indicates significance at 95% level; ** indicates significance at 90% level.

6.1 Probabilities

We do not prove the following here; but it is true.

$$P_{ij} = \frac{e^{\beta' X_{ij} \; + \; \psi'_j A_i}}{\sum_{k=1}^{J} e^{\beta' X_{ik} \; + \; \psi'_k A_i}}$$

$$= \frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}}$$

Same probabilities as MNL.

Just another form of MNL.

7 Goodness of Fit, Predicted Values, Classification

There is no R^2.

"Pseudo-R^2" $\rightarrow$ even worse than R^2.

Compute $\hat{P}_{ij}$.

$$\hat{P}_{ij} \rightarrow \hat{Y}_i$$

Classification Rule (binomial):

$$\hat{Y}_i = 1 \text{ if } \hat{P}_i > .5$$

Percent Correctly Predicted:

"correct prediction": $\hat{Y}_i = 1$ and $Y_i = 1$ or, $\hat{Y}_i = 0$ and $Y_i = 0$.

$$\textbf{PCP} = \frac{100(\#\ \textbf{of Correct Predictions})}{N}$$

$$\textbf{PMC} = \textbf{Percent in Modal Category}$$

7.1 PRE [Proportional Reduction in Error]:

$$\textbf{PRE} = \frac{\textbf{PCP} - \textbf{PMC}}{1 - \textbf{PMC}}$$

Example 1:

PCP = .85 PMC = .80

$$\textbf{PRE} = \frac{.85 - .80}{1 - .80} = \frac{.05}{.20} = .25$$

Example 2:

PCP = .75 PMC = .50

$$\textbf{PRE} = \frac{.75 - .50}{1 - .50} = \frac{.25}{.50} = .50$$

Classification Rule (multinomial):

If you require $\hat{P}_{ij} > .5$ for $\hat{y}_i = j$, then you may not classify some observations.

So:

$\hat{y}_i = j \quad \text{if} \quad \hat{P}_{ij} > \hat{P}_{ik} \quad \forall k \neq j.$

7.2 Unevenly Distributed Data:

If the data has a very skewed distribution (90% 0's; 10% 1's), you may **never** predict 1 as an outcome.

Common in multinomial cases where one case may be relatively rare.

A Useful Table (hypothetical numbers):

		Observed			
		Cons	**Labour**	**All**	$\widehat{\textbf{Total}}$
	$\widehat{\textbf{Cons}}$	300	25	10	335
Pred	$\widehat{\textbf{Labour}}$	5	250	65	320
	$\widehat{\textbf{Alliance}}$	15	10	100	125
	Total	320	285	175	780

This table has all the information (except uncertainty).

We can see that we do not predict votes for Alliance very well.

7.3 cPCP (Herron - 1999)

Problem: We treat $\hat{P}_i = .51$ and $Y_i = 1$ the same as $\hat{P}_i = .95$ and $Y_i = 1$. But, we should give more credit for the latter prediction: it is a 'better' prediction.

Solution: "Expected PCP".

$$\mathbf{ePCP} = \frac{1}{N}\left(\sum_{Y_i=1}\hat{P}_i + \sum_{Y_i=0}(1 - \hat{P}_i)\right)$$

Example:

Y_i	$\hat{P}_i$
0	.6
1	.6
1	.8

ePCP = 1/3 (.4 + .6 + .8) = .6

7.4 ePCP - Multinomial

$$\mathbf{ePCP} \;=\; \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J}\left(\hat{P}_{ij}\cdot(y_i == j)\right)$$

$$\mathbf{ePCP} \;=\; \frac{1}{3}\left(\sum_{Y_i=1}\hat{P}_{i1} \;+\; \sum_{Y_i=2}(\hat{P}_{i2}) \;+\; \sum_{Y_i=3}(\hat{P}_{i3})\right)$$

8 First Differences

$$P_i = \Phi(\beta' X_i) \tag{24}$$

Say: we want to know what happens if X_{ik} were to change to $\tilde{X}_{ik}$; say X_{ik} were to increase by z units.

[X_{ik} is a particular element of the vector X_i.]

1. Compute: $\hat{P}_i \ = \ \Phi(\beta' X_i)$
2. Compute: $\tilde{X}_{ik} \ = \ X_{ik} \ + \ z$
3. Compute: $\tilde{V}_i \ = \ \beta' \tilde{X}_i$
4. Compute: $\tilde{P}_i \ = \ \Phi(\tilde{V}_i)$
5. Compute: $\tilde{P}_i \ - \ \hat{P}_i$

The last difference is the quantity of interest.

We could also compute this for **everyone** in the sample, and then compute the mean of $\tilde{P}_i \ - \ \hat{P}_i$; and get the effect of **all** respondents changing their taste on characteristic k by z units.

9 Computing Effects of Changes in Characteristics of a Respondent

$$P_{ij} = \frac{e^{\beta' X_{ij} \ + \ \psi'_j A_i}}{\sum_{k=1}^{J} e^{\beta' X_{ik} \ + \ \psi'_k A_i}} \tag{25}$$

Say: we want to know what happens if a_i were to change to $\tilde{a}_i$; say a_i were to increase by z units. [a_i is a particular element of the vector A_i.]

1. Compute : $\tilde{a}_i \ = \ a_i \ + \ z$
2. Compute: $\tilde{V}_{ij} \ = \ \beta' X_{ij} \ + \ \psi'_j \tilde{A}_i$
3. Compute $\tilde{P}_{ij}$
4. Compute: $\tilde{P}_{ij} \ - \ \hat{P}_{ij}$

The last difference is the quantity of interest.

We could also compute this for **everyone** in the sample, and then compute the mean of $\tilde{P}_{ij} \ - \ \hat{P}_{ij}$; and get the effect of **all** respondents changing their taste on characteristic a by z units.

Table 4 (*A/N 1995 - AJPS*)
Effects of Economics, Issues, and Anger in the 1992 Election

		Probability of Voting For:		
		Bush	Clinton	Perot
Personal Finances	Better	0.42	0.31	0.27
	Worse	0.35	0.35	0.29
	Difference	0.07	-0.05	-0.02
National Economy	Better	0.54	0.19	0.27
	Worse	0.24	0.49	0.27
	Difference	0.29	-0.30	0.00
Voter Ideology[a]	Near	0.46	0.39	0.31
	Far	0.32	0.28	0.20
	Difference	0.14	0.11	0.12
Minorities	Assist.	0.30	0.48	0.22
	No Assist.	0.46	0.20	0.33
	Difference	-0.17	0.27	-0.11
Abortion	Pro-Life	0.62	0.22	0.16
	Pro-Choice	0.28	0.38	0.34
	Difference	0.34	-0.16	-0.18
Term Limits	For	0.39	0.33	0.28
	Against	0.38	0.32	0.30
	Difference	0.01	0.01	-0.02

Note: Table entries are the predicted probabilities of a hypothetical individual voting for Clinton, Bush or Perot based on different values of the row-variable.

[a] Probabilities for each of the candidates in the voter-ideology row are based on the ideological distance between the voter and the particular candidate.

10 Computing Effects of Changes in Characteristics of an Alternative

$$\hat{P}_{ij} = \frac{e^{\hat{\beta}' X_{ij} \; + \; \hat{\psi}'_j A_i}}{\sum_{k=1}^{J} e^{\hat{\beta}' X_{ik} \; + \; \hat{\psi}'_k A_i}} \tag{26}$$

We want to alter A_i, and recompute $\hat{P}_{ij}$.

Say:

$$X_{ij} \; = \; (\text{resp}_i \; - \; \text{choice}_j) \tag{27}$$

We want to know what happens if the j^{th} choices 'moves' z units to the right.

1. Set: $\widetilde{\text{choice}}_j \; = \; \text{choice}_j \; + \; z$
2. Compute: $\tilde{X}_{ij} \; = \; (\text{resp}_i \; - \; \widetilde{\text{choice}}_j)$
3. Compute: $\tilde{V}_{ij} \; = \; \beta' \tilde{X}_{ij} \; + \; \psi'_j A_i$
4. Compute $\tilde{P}_{ij}$
5. Compute: $\tilde{P}_{ij} \; - \; \hat{P}_{ij}$

The last difference is the quantity of interest.

Table 5 (*A/N 1998 - AJPS*)
Conditional Logit Estimates of Effect of Movement By the Labour Party +/- $\frac{1}{2}$ Standard Deviation - British Election - 1987

		Conservatives	Labour	Alliance
Baseline		45.2	29.5	25.3
Defense	— $\frac{1}{2}\sigma$	45.7	28.3	26.0
	+ $\frac{1}{2}\sigma$	44.7	30.6	24.8
	Difference	-1.0	2.3	-1.3
Phillips	— $\frac{1}{2}\sigma$	45.2	29.7	25.2
	+ $\frac{1}{2}\sigma$	45.3	29.0	25.7
	Difference	0.2	-0.7	0.6
Taxation	— $\frac{1}{2}\sigma$	45.6	28.6	25.8
	+ $\frac{1}{2}\sigma$	45.1	29.4	25.5
	Difference	-0.5	0.8	-0.3
Nationalization	— $\frac{1}{2}\sigma$	45.9	27.7	26.4
	+ $\frac{1}{2}\sigma$	44.6	30.8	24.6
	Difference	-1.3	3.1	-1.8
All Issues	— $\frac{1}{2}\sigma$	47.1	24.9	28.0
	+ $\frac{1}{2}\sigma$	43.5	31.8	24.7
	Difference	-3.6	6.8	-3.2

Note: Estimated impact of the Labour party moving from one half a standard deviation to the left of its mean perceived position to one half a standard deviation to the right of its mean perceived position on each of seven issues. Column entries are estimated aggregate vote-shares.

11 To Compute Confidence Intervals:

1. Estimate the model.
2. This produces estimates of the parameters (β), and the variance-covariance matrix of the parameters (V).
3. Draw a set of values of the parameters from a multivariate normal distribution: $N(\hat{\beta}, \hat{V})$.
4. Compute predicted values of the quantity of interest for: 1) a specified respondent or case; or 2) the entire sample; or 3) some relevant subsample.
5. Repeat the above procedure N times where N is a big number (500 or 1000 is generally realistic).
6. The sampling distribution of the predicted probabilities will give you the confidence interval.

See "Clarrify"; URL = http://gking.harvard.edu

12 Independence of Irrelevant Alternatives (IIA)

$$\frac{P_{ij}|S_s}{P_{ik}|S_s} = \frac{P_{ij}|S_p}{P_{ik}|S_p} \quad \forall \quad j,k,s,p \tag{28}$$

where:

- S_s and S_p denote sets of alternatives
- $j,k \ \in \ S_p$, and $j,k \ \in \ S_s$
- and $P_{ij}|S_s$ denotes the probability of the i^{th} individual choosing alternative j from choice-set S_s.

Notice: There is no statement about econometrics or disturbance terms on this page!

Remember the Conditional Logit and Multinomial Logit Probabilities:

$$P_{ij} \;=\; \frac{e^{V_{ij}}}{\sum_{t=1}^{J} e^{V_{it}}}$$

Notice:

$$\frac{P_{ij}}{P_{ik}} \;=\; \frac{e^{V_{ij}}}{e^{V_{ik}}}$$

Table 1

Characteristics of Discrete Choice Models

	Multinomial Logit	Conditional Logit	Generalized Extreme Value	Multinomial Probit
Alternative Specific Variables	No	Yes	Yes	Yes
Correlated Disturbances	No	No	Some	Yes
Includes Position of Party	No	Yes	Yes	Yes
Can Correctly Measure Movement by Parties	No	Yes	Yes	Yes
Assumes IIA	Yes	Yes	No	No
Can Correctly Measure Omission of a Party	No	No	Sometimes	Yes

13 Multinomial Probit

Define a random utility function for voter i over each candidate j:

$$U_{ij} \;=\; \beta\,\mathbf{X_{ij}} \;+\; \psi_j \mathbf{A_i} + \varepsilon_{ij}$$

Assume that the random elements of the utility functions, ε_{ij}, have a multivariate normal distribution with a mean vector zero and a covariance matrix:

$$\Sigma_i = \begin{bmatrix} \sigma^2_{i,1} & & \\ \sigma_{i,12} & \sigma^2_{i,2} & \\ \sigma_{i,13} & \sigma_{i,23} & \sigma^2_{i,3} \end{bmatrix}$$

$$\begin{aligned} P_{i1} = & \quad Pr[(U_{i1} > U_{i2}) \quad \& \quad (U_{i1} > U_{i3})] \\ = & \quad Pr[(V_{i1} + \varepsilon_{i1} > V_{i2} + \varepsilon_{i2}) \quad \& \quad (V_{i1} + \varepsilon_{i1} > V_{i3} + \varepsilon_{i3})] \\ = & \quad Pr[(\varepsilon_{i2} - \varepsilon_{i1} \; < \; V_{i1} - V_{i2}) \quad \& \quad (\varepsilon_{i3} - \varepsilon_{i1} \; < \; V_{i1} - V_{i3})] \end{aligned}$$

The insight of Hausman and Wise (1978) was to let

$$\begin{aligned} \eta_{i,21} &= \varepsilon_{i2} - \varepsilon_{i1}, \\ \eta_{i,31} &= \varepsilon_{i3} - \varepsilon_{i1}. \end{aligned}$$

The joint distribution for the $\eta_{i,j1}$ will be bivariate normal, which has a covariance matrix:

$$\Omega_{i1} \;=\; \begin{bmatrix} \sigma^2_{i,1} + \sigma^2_{i,2} - 2\sigma_{i,12} & \\ \sigma^2_{i,1} - \sigma_{i,13} - \sigma_{i,12} + \sigma_{i,23} & \sigma^2_{i,1} + \sigma^2_{i,3} - 2\sigma_{i,13} \end{bmatrix}$$

This allows us to write the probability that voter i will choose candidate 1 as:

$$P_{i1} = \int_{-\infty}^{\frac{V_{i1}-V_{i2}}{\sqrt{\sigma^2_{i,1}+\sigma^2_{i,2}-2\sigma_{i,12}}}} \int_{-\infty}^{\frac{V_{i1}-V_{i3}}{\sqrt{\sigma^2_{i,1}+\sigma^2_{i,3}-2\sigma_{i,13}}}} b_1(\eta_{i,21}, \eta_{i,31}; r_1) d\eta_{i,21} d\eta_{i,31}$$

with b_1 being the standardized bivariate normal distribution and r_1 being the correlation between $\eta_{i,21}$ and $\eta_{i,31}$.

The log-likelihood is given by:

$$L = K + \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} log P_{ij}$$

where:

$$\frac{\partial L}{\partial \beta_k} = \sum_{i=1}^{N} \sum_{j=1}^{J} \;\; \frac{y_{ij}}{P_{ij}} \;\; \frac{\partial P_{ij}}{\partial \beta_k}$$

14 Issues With MNP

Identification: We normalize $\sigma_j^2 = 1$, and we can estimate up to $\frac{J(J-1)}{2} - 1$ covariance terms of Σ; assume the others are 0.

The choice of which elements of Σ to set to 0 is just a normalization, it does not matter which ones you set to 0.

But: you must estimate σ_{ij} such that: $-1 < \sigma_{ij}$¡1.

Interpretation of Σ: ????

Fragile Identification: (Keane)

The likelihood can be fairly flat.

It may help to omit some Xs from the utility of some alternatives.

Sample Size: More is better. Probably trouble with samples below 750.

Results: MNP and CL results will differ most when simulating the effect of removing an alternative from the choice set.

14.1 Hausman/McFadden Specification Test for IIA:

1) Estimate the full model with CL (or MNL): this produces $\hat{b}eta_f$ and $\hat{V}_f$.

2) Omit k choices, estimate the model on the remaining subset of choices using CL (or MNL): this produces $\hat{\beta}_s$ and $\hat{V}_s$.

3) The following test statistic is distributed as a χ^2 random variable with k degrees of freedom:

$$\chi^2 = (\hat{\beta}_s - \hat{\beta}_f)'[\hat{V}_s - \hat{V}_f]^{-1}(\hat{\beta}_s - \hat{\beta}_f)$$

This is an easy test to do!!

14.2 IIA Does Not Aggregate (*A/N 1998 - Appendix C*)

Assume the i^{th} voter's utility for the j^{th} party is given by:

$$U_{ij} = -(x_i - C_j)^2 + \epsilon_{ij} \tag{29}$$

where:

- x_i is the i^{th} voter's position
- C_j is the j^{th} party's position
- ϵ_{ij} is a random disturbance term with an extreme value distribution

Assume we initially have two parties:

- Liberal (L) at -1
- Conservative (C) at 1

Add a third party (Moderates - M) at .5.

Assume we have a five person electorate with voters at: -2, -.5, 0, .25, and .75.

If one looks at the ratio of

$$\frac{P_{iL}|L,C}{P_{iC}|L,C}$$

and compares it to

$$\frac{P_{iL}|L,C,M}{P_{iC}|L,C,M}$$

for any respondent, they are equal.

However, if one looks at the ratio of the *means* of P_{iL} to P_{iC} across the two different hypothetical elections they are different. In the first race the Conservatives are predicted to have 46% of the two way vote. In the three way race the Conservatives would only have 37% (i.e., .23 / (.23 + .40)) of the two way vote between the Conservatives and Liberals. Thus, consistent with our intuition, the entry of the Right-Moderates — a second right party — takes more votes from the Conservatives than from the Liberals.

Table B3

Vote Shares and Individual Probabilities: IIA Does Not Aggregate

	Probabilities From Two Cand Race		Probabilities From Three Cand Race		
V's Pos	$P_C\|\{CL\}$	$P_L\|\{CL\}$	$P_C\|\{CLM\}$	$P_L\|\{CLM\}$	$P_M\|\{CLM\}$
-2	.00	1.00	.00	.99	.01
-.5	.12	.88	.08	.62	.29
0	.50	.50	.24	.24	.51
.25	.73	.27	.33	.12	.55
.75	.95	.05	.49	.02	.49
Mean	.46	.54	.23	.40	.37

15 Multinomial Logit is Equivalent to Binomial Logit

(*A/N 1998 - Appendix B*)

$$
\begin{aligned}
U_{i1} &= \beta_1' X_i + \epsilon_{i1} \\
U_{i2} &= \beta_2' X_i + \epsilon_{i2} \\
U_{i3} &= \beta_3 X_i + \epsilon_{i3}
\end{aligned}
$$

β_j represents a parameter vector determining the contribution of voter characteristics to utility for choice j.

In both multinomial logit and binomial logit one parameter is normalized to zero. Thus, when estimating binomial logit only one set of coefficients is produced since the other set has been normalized to zero and never appears.

In this case, say β_3 is normalized to zero. Then if choice 1 is omitted, binomial logit will generate consistent estimates of β_2 using only individuals who pick choice 2 or choice 3.

Omitting choice 2, binomial logit will generate consistent estimates of β_1.

16 MNL as Reduced Form of CL

We first define several things:

- $X_i \;=\; i^{th}$ Voter's Position in the issue space
- $C_j \;=\; j^{th}$ Party's Position in the Issue Space
- $D_{ij} \;=\;$ Distance from i^{th} Voter to j^{th} Party

Now according to the classic spatial model; the i^{th} individual's utility of the j^{th} choice is:

$$V_{ij} = -\beta_j \;*\; (X_i \;-\; C_j)^2 \tag{30}$$

or,

$$V_{ij} = -\beta_j \;*\; D_{ij} \tag{31}$$

This fits nicely into the conditional logit random utility model (RUM) setup:

$$U_{ij} = -\beta_j \;*\; D_{ij} \;+\; u_i \tag{32}$$

Notice that:

$$\begin{aligned} D_{i1} &= X_i^2 - 2X_iC_1 + C_1^2 \\ D_{i2} &= X_i^2 - 2X_iC_2 + C_2^2 \\ D_{i1} - D_{i2} &= -2X_i(C_1 - C_2) + (C_1^2 - C_2^2) \\ &= -2xa^* + b^* \end{aligned}$$

where

$$\begin{aligned} a^* &= (C_1 - C_2) \\ b^* &= (C_1^2 - C_2^2) \end{aligned}$$

The difference between the two distances is a linear function of the voter's position. This suggests that if:

$$Pr(Y_i = 1) = f(\alpha_o + \beta_1(D_{i1} - D_{i2}))$$

and α_0 and β_0 are identified then we can substitute $-2X_ia^* + b^*$ for $D_{i1} - D_{i2}$ and recover reduced form estimates:

$$\begin{aligned} Pr(Y_i = 1) &= f(\alpha_o + \beta_1(-2X_ia^*) + \beta_1 b^*) \\ &= f((\alpha_0 + \beta_1 b^*) + (-2a^*\beta_1)X_i) \end{aligned}$$

So one could recover:

$$\begin{aligned} \tilde{\alpha} &= \alpha_0 + \beta_1 b^* \\ \tilde{\beta} &= -2a^*\beta_1 \end{aligned}$$

from standard MNL estimates (i.e.; binomial logit in this case). Again, note that this result holds assuming quadratic utility functions.

17 The GEV Model

We assume the utility of the i^{th} individual for the j^{th} choice is given as follows:

$$U_{ij} = X_{ij}\beta \ + \ a_i\psi_j \ + \ \epsilon_{ij}$$

where: ϵ_{ij} is a disturbance term with an extreme value distribution given below.

Extreme Value Distribution:

$$F(\epsilon_1, \epsilon_2, \epsilon_3) = exp[-G(e^{-\epsilon_1}, e^{-\epsilon_2}, e^{-\epsilon_3})]$$

where G is a nonnegative function with the constraints that it be homogeneous of degree 1 and always greater than or equal to 0.

For cases with three alternatives where the operative hypothesis is that alternatives 2 and 3 are grouped, G can be defined as follows:

$$G(Y_1, Y_2, Y_3) \ = \ Y_1 \ + \ (Y_2^{1/(1-\sigma)} \ + \ \ Y_3^{1/(1-\sigma)})^{(1-\sigma)}$$

σ is a parameter to be estimated, and is commonly interpreted as a measure of the correlation between the stochastic elements of alternatives 2 and 3.

Furthermore, σ is bounded such that $0 \leq \sigma < 1$. Thus an estimated value of $\hat{\sigma}$ outside these bounds suggests a misspecification problem with the model: the systemic component could be misspecified, or the grouping could be misspecified, or both. This leads to the hope that an estimated σ within the range $(0, 1]$ suggests a correct model specification.

Following Maddala (1983) we simplify the notation by letting $Y_{ij} = e^{V_{ij}}$, where V_{ij} denotes the systemic component of utility:

$$V_{ij} = a_i \psi_j + X_{ij} \beta$$

Probabilities for the generalized extreme-value model are given by:

$$\begin{aligned}
P_{i1} &= \frac{Y_{i1}}{G(Y_{i1}, Y_{i2}, Y_{i3})} \\
P_{i2} &= \frac{Y_{i2}^{1/(1-\sigma)}(Y_{i2}^{1/(1-\sigma)} \ + \ Y_{i3}^{1/(1-\sigma)})^{-\sigma}}{G(Y_{i1}, Y_{i2}, Y_{i3})} \\
P_{i3} &= \frac{Y_{i3}^{1/(1-\sigma)}(Y_{i2}^{1/(1-\sigma)} \ + \ Y_{i3}^{1/(1-\sigma)})^{-\sigma}}{G(Y_{i1}, Y_{i2}, Y_{i3})}
\end{aligned}$$

The log-likelihood function is:

$$LL = \Sigma(y_1 ln(P_{i1}) \ + \ y_2 ln(P_{i2}) \ + \ y_3 ln(P_{i3}))$$

This is straightforward to evaluate with:

$$\begin{aligned}
ln(P_1) &= V_1 - ln(G) \\
ln(P_2) &= \left(\frac{1}{1-\sigma}\right) V_2 \ - \ \sigma ln(e^{V_2^{\frac{1}{1-\sigma}}} \ + \ e^{V_2^{\frac{1}{1-\sigma}}}) \ - \ ln(G) \\
ln(P_3) &= \left(\frac{1}{1-\sigma}\right) V_3 \ - \ \sigma ln(e^{V_2^{\frac{1}{1-\sigma}}} \ + \ e^{V_2^{\frac{1}{1-\sigma}}}) \ - \ ln(G)
\end{aligned}$$

18 Heteroscedastic Probit

Note: In standard probit and logit, the estimators are not consistent in the presence of heteroscedasticity.

We have the usual latent variable setup:

$$\begin{aligned} y_i^* &= X_i\beta + \epsilon_i \\ y &= 1 \; if \; y^* > 0 \\ y &= 0 \; if \; y^* \leq 0 \end{aligned}$$

But, we assume that:

$$\mathbf{Var}(\epsilon_i) = \left(e^{z_i\gamma}\right)^2$$

$$\begin{aligned} \mathbf{Pr}(Y_i = 1) &= \mathbf{Pr}(y_i^* > 0) \\ &= \mathbf{Pr}(X_i\beta + \epsilon_i > 0) \\ &= \mathbf{Pr}(\epsilon_i > -X_i\beta) \\ &= 1 - \mathbf{Pr}(\epsilon_i < -X_i\beta) \end{aligned}$$

By Assumption:

$$\epsilon_i \sim \mathbf{N}\left(0, \left(e^{z_i\gamma}\right)^2\right)$$

Since F is symettric about 0:

$$\begin{aligned}\mathbf{Pr}(Y_i \ = \ 1) \ &= \ \mathbf{Pr}(\epsilon_i \ < \ X_i\beta) \\ &= \ \Phi\left(\frac{X_i\beta}{e^{z_i\gamma}}\right)\end{aligned}$$

$$L \ = \ \prod_{y_i=1} \Phi\left(\frac{X_i\beta}{e^{z_i\gamma}}\right) \ \cdot \ \prod_{y_i=0}\left(1-\Phi\left(\frac{X_i\beta}{e^{z_i\gamma}}\right)\right)$$

$$\begin{aligned}\textbf{Log L} \ &= \ \sum y_i \textbf{log}\ \Phi\left(\frac{X_i\beta}{e^{z_i\gamma}}\right) \\ &\quad + \ \sum(1 \ - \ y_i)\textbf{log}\left[1 \ - \ \Phi\left(\frac{X_i\beta}{e^{z_i\gamma}}\right)\right]\end{aligned}$$

Test for Homoscedasticity:

$$\mathbf{LR} \ = \ 2\left(\mathbf{LL_{HPROB}} \ - \ \mathbf{LL_{PROB}}\right) \ \sim \ \chi^2_k$$

where k is the number of parameters in γ.

19 Nested Multinomial Logit (McFadden - 1981; Maddala)

Housing Choice:

Communities: i = 1,2,..., C
Dwellings : j = 1,2, ..., N_i in community i

$$U_{ij} \; = \; V_{ij} \; + \; \epsilon_{ij}$$

If ϵ_{ij} are iid, extreme value:

$$\mathbf{P_{ij}} \; = \; \frac{e^{V_{ij}}}{\sum_{m=1}^{C} \sum_{n=1}^{N_m} \; e^{V_{mn}}}$$

This is just conditional logit with different subscripts.

$$V_{ij} \; = \; \beta' X_{ij} \; + \; \alpha' Y_i$$

X_{ij} varies with community and dwelling.
Y_i varies only by community.

$P_{ij} \; = \; P_{j|i} \; \cdot \; P_i$

Probability of choosing j^{th} dwelling in i^{th} community.

$$
\begin{aligned}
P_{j|i} &= \frac{e^{V_{ij}}}{\sum_{k=1}^{N_i} e^{V_{ik}}} \\
&= \frac{e^{\beta' X_{ij} \ + \ \alpha' Y_i}}{\sum_{k=1}^{N_i} e^{\beta' X_{ik} \ + \ \alpha' Y_i}} \\
&= \frac{e^{\beta' X_{ij}}}{\sum_{k=1}^{N_i} e^{\beta' X_{ik}}}
\end{aligned}
$$

The Y_i dropped out!

$$
\begin{aligned}
P_i &= \frac{\sum_{j=1}^{N_i} e^{V_{ij}}}{\sum_{m=1}^{C} \sum_{n=1}^{N_m} \ e^{V_{mn}}} \\
&= \frac{\sum_{j=1}^{N_i} e^{\beta' X_{ij} \ + \ \alpha' Y_i}}{\sum_{m=1}^{C} \sum_{n=1}^{N_m} \ e^{\beta' X_{mn} \ + \ \alpha' Y_i}}
\end{aligned}
$$

Following is (3.11):

$$
P_i = \frac{e^{\alpha' Y_i} \sum_{j=1}^{N_i} e^{\beta' X_{ij}}}{\sum_{m=1}^{C} \left(e^{\alpha' Y_m} \ \left(\sum_{n=1}^{N_m} e^{\beta' X_{mn}} \right) \right)}
$$

Now Define the 'Inclusive Value' for all houses in community i:

$$I_i \;=\; log\left(\sum_{j=1}^{N_i} e^{\beta' X_{ij}}\right)$$

Now (3.13) by substituting I into (3.10) and then (3.14) by substituting I into (3.11):

$$(3.13) \quad P_{j|i} \;=\; \frac{e^{\beta' X_{ij}}}{e^{I_i}}$$

$$(3.14) \quad P_i \;=\; \frac{e^{\alpha' Y_i \;+\; I_i}}{\sum_{m=1}^{C} e^{\alpha' Y_m \;+\; I_m}}$$

To Estimate:

1) Estimate β from the conditional choice model ($P_{j|i}$ in equation 3.10.

2) Compute I: a fuction of β and X.

3) Estimate α from the equation for choosing a community (particular group) i in equation 3.14. Notice that α is only a function of I and Y.

Repeat (3.14):

$$(3.14) \quad P_i \; = \; \frac{e^{\alpha' Y_i \; + \; I_i}}{\sum_{m=1}^{C} e^{\alpha' Y_m \; + \; I_m}}$$

Now let the coefficient on I take a value $\neq 1$.

$$(3.23) \quad P_i \; = \; \frac{e^{\alpha' Y_i \; + \; (1-\sigma) I_i}}{\sum_{m=1}^{C} e^{\alpha' Y_m \; + \; (1-\sigma) I_m}}$$

(3.23) is "nested logit".

N alternatives

M groups

N_i alternatives in i^{th} group, so $\sum_m N_i = N$.

$$G(Y) = \sum_{i=1}^{M} a_i \left(\sum_{j=1}^{N_i} Y_{ij}^{\frac{1}{1-\sigma_i}} \right)^{1-\sigma_i}$$

$a_i > 0; 0 \leq \sigma_i < 1$

$Y_{ij} = e^{V_{ij}}$

19.1 Born

3 choices:

1) absention
2) in party
3) out party

Group: (in-party, out-party)

$$P_{i1} \;=\; \frac{e^{V_{i1}}}{e^{V_{i1}} \;+\; \left(e^{V_{i2}/d} \;+\; e^{V_{i3}/d}\right)^{d}}$$