

Categorical Models

Brad Jones¹

¹Department of Political Science
University of California, Davis

April 28, 2009

Ordinal Response Variables

- ▶ Definition: Confine ourselves to y well understood to be limited (3-7 points scales, for example)
- ▶ Basic problems emerge but are often ignored.
- ▶ Regression is commonly applied:
“Because the dependent variables are categorical, OLS regression is technically inappropriate. We found substantially the same results, however, using ordinal logit models. We report the OLS results because their interpretation is more straightforward.” (Zuckerman and Jost, 2001 SPQ).

Illustration

- ▶ Suppose we proceed with OLS?
- ▶ Response Variable: 4-point Likert scale on attitudes toward immigration.
- ▶ Use same covariates from last week.
- ▶ Model returns:

OLS Estimates for Immigration Effect Variable

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.4332	0.1490	16.33	0.0000
pid	-0.3022	0.0604	-5.01	0.0000
education	0.0136	0.0215	0.63	0.5280
income	0.1049	0.0379	2.77	0.0059
latino	-0.8992	0.1088	-8.26	0.0000

Problems

- ▶ Equal-interval scoring assumed.
- ▶ “a unit change in x results in a β change in $E(Y)$ ”
- ▶ False sense of precision.
- ▶ Alternative models need to be considered (or should be!)

Utility Motivation

- ▶ Assume Y is a discretized measure of Y^*
- ▶ Postulate

$$Y_i^* = \alpha + \mathbf{x}'\beta + \epsilon \quad (1)$$

- ▶ Cut Point Rule:

$$Y = 1 \quad \equiv \quad Y^* \leq \alpha_1$$

$$Y = 2 \quad \equiv \quad \alpha_1 < Y^* \leq \alpha_2$$

$$Y = 3 \quad \equiv \quad \alpha_2 < Y^* \leq \alpha_3$$

$$Y = 4 \quad \equiv \quad Y^* > \alpha_3$$

Utility Motivation

- ▶ Easy to visualize on a line.
- ▶ The idea is to map the latent variable Y^* onto Y .
- ▶ As in previous slides, we identify thresholds or cutpoints that partition the space.
- ▶ Since these are unknown, we estimate them given the observed data.
- ▶ With likelihood, this means we need to identify some function $F(\cdot)$.
- ▶ That is, a distribution function suitable for ordinal categorical data.
- ▶ Equivalently, specify a distribution function for ϵ in equation (1).

Probability

- ▶ Getting to some data.
- ▶ Y is a j -category variable.
- ▶ In terms of cumulative probabilities:

$$C_{ij} = \Pr(y_i \leq j) = \sum_{k=1}^j \Pr(y_i = k) \quad (2)$$

- ▶ The cumulative probabilities must sum to 1.
- ▶ Because of this, we have to impose a constraint on the model:
- ▶ Only $J - 1$ cumulative probabilities are uniquely identified.
- ▶ Thus, only $J - 1$ thresholds need to be estimated.

Probability

- ▶ Define a model:

$$C_{ij} = F(\alpha_j + \mathbf{x}'\beta) \quad (3)$$

- ▶ Conditional Probabilities

$$\Pr(y_i | \mathbf{x}) = \begin{cases} F(\alpha_1 + \mathbf{x}'\beta) & j = 1, \\ F(\alpha_2 + \mathbf{x}'\beta) - F(\alpha_1 + \mathbf{x}'\beta) & 1 < j \leq 2, \\ F(\alpha_3 + \mathbf{x}'\beta) - F(\alpha_2 + \mathbf{x}'\beta) & 2 < j \leq 3, \\ 1 - F(\alpha_3 + \mathbf{x}'\beta) & j = J. \end{cases}$$

- ▶ Looks similar to the cutpoint rule from before.
- ▶ In fact, it is identical.
- ▶ Note something critically important in the relationships shown above.
- ▶ The effects of the covariates are invariant to the cutpoints.
- ▶ Means the cutpoints have the effect of shifting the probability curve to the left or right.

Ordinal Models

- ▶ Probability curves will be identical; just shifted along the x -axis.
- ▶ Generically, this is called the “parallel” slopes assumptions.
- ▶ In the logit setting, it is called the *proportional odds* assumption.
- ▶ To test this assumption, we first need to know how to estimate a model!
- ▶ A couple of good candidates for $F(\cdot)$ are logistic distribution and standard normal.

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x)^2\right) dx \quad (4)$$

- ▶ Under probit we get:

$$\Pr(y_i = j \mid x) = F(\alpha_{j-1} \leq \mathbf{x}'\boldsymbol{\beta} \leq \alpha_j) = \Phi(\alpha_{j-1} \leq \mathbf{x}'\boldsymbol{\beta} \leq \alpha_j) \quad (5)$$

- ▶ Φ is the cdf for the standard normal

Ordinal Models: Ordinal Probit

- ▶ I've written the probit function using the "latent utility" motivation.
- ▶ It need not be written in this way (see point 2 from two slides back).
- ▶ Sometimes α will be referenced as τ .
- ▶ Cumulative probabilities in probit.
- ▶ Let Z be the probit linear prediction: $Z_j = \sum_{k=1}^K \mathbf{x}'\beta$.

$$\Pr(y_i = 1 \mid \mathbf{x}) = \Phi(\alpha_1 - Z_j)$$

$$\Pr(y_i = 2 \mid \mathbf{x}) = \Phi(\alpha_2 - Z_j) - \Phi(\alpha_1 - Z_j)$$

$$\Pr(y_i = 3 \mid \mathbf{x}) = \Phi(\alpha_3 - Z_j) - \Phi(\alpha_2 - Z_j)$$

$$\Pr(y_i = 4 \mid \mathbf{x}) = 1 - \Phi(\alpha_3 - Z_j).$$

- ▶ These probabilities fully summarize the probability space for a 4-category response variable.

Ordinal Models: Ordinal Logit

- ▶ Retrieving the ordinal logit means specifying $F(\cdot)$ in terms of the logistic.

$$\Pr(Y \leq y_j | \mathbf{x}) = \frac{\exp(\alpha_j - \mathbf{x}'\beta)}{1 + \exp(\alpha_j - \mathbf{x}'\beta)}$$

- ▶ Linear Model for log-odds:

$$\log \left[\frac{\Pr(Y \leq y_j | \mathbf{x})}{\Pr(Y > y_j | \mathbf{x})} \right] = \alpha_j - \mathbf{x}'\beta, \quad j = 1, 2, \dots, j-1$$

- ▶ Proportional Odds Property:

$$\frac{\exp(x_1\beta)}{\exp(x_2\beta)} = \exp\{(x_1 - x_2)'\beta\}$$

Ordinal Models: Ordinal Logit

- ▶ Conditional Probabilities (letting Z be linear predictor):

$$\Pr(y_i = 1 \mid \mathbf{x}) = \frac{1}{1 + \exp(Z_j - \alpha_1)}$$

$$\Pr(y_i = 2 \mid \mathbf{x}) = \frac{1}{1 + \exp(Z_j - \alpha_2)} - \frac{1}{1 + \exp(Z_j - \alpha_1)}$$

$$\Pr(y_i = 3 \mid \mathbf{x}) = \frac{1}{1 + \exp(Z_j - \alpha_3)} - \frac{1}{1 + \exp(Z_j - \alpha_2)}$$

$$\Pr(y_i = 4 \mid \mathbf{x}) = 1 - \frac{1}{1 + \exp(Z_j - \alpha_3)}$$

(6)

- ▶ As with binary logit/probit, the ordinal versions will be very similar.
- ▶ Virtually indistinguishable from one another in many cases.

Estimation

- ▶ Estimation through standard likelihood-based methods.
- ▶ Function is generally well behaved; Newton-Raphson is commonly hardcoded/so is Fisher scoring.
- ▶ General log-likelihood form:

$$\log L = \sum_{i=1}^n \sum_{j=1}^J d_{ij} \log[F(\alpha_j - \mathbf{x}'\beta) - F(\alpha_{j-1} - \mathbf{x}'\beta)] \quad (7)$$

- ▶ Estimation in R using the MASS package and polr option/ Stata's ologit or oprobit works.
- ▶ Turn to example using immigration data. First logit.

```
> ologit.mod<-polr(y ~ pid + education +
+ income + latino, method = c("logistic"))
```

```
> summary(ologit.mod)
```

Re-fitting to get Hessian

Call:

```
polr(formula = y ~ pid + education + income + latino, method = c("logistic"))
```

Coefficients:

	Value	Std. Error	t value
pid	-0.58236556	0.11984650	-4.8592621
education	0.02912987	0.04249420	0.6855021
income	0.19449400	0.07509927	2.5898254
latino	-1.87599413	0.23360995	-8.0304547

Intercepts:

	Value	Std. Error	t value
1 2	-1.3725	0.3082	-4.4532
2 3	0.2447	0.2984	0.8203
3 4	1.5395	0.3026	5.0878

Residual Deviance: 1104.433

AIC: 1118.433

(6 observations deleted due to missingness)

```
>
```

Ordinal Models: Ordinal Logit

- ▶ Coefficients are log-odds ratios.
- ▶ Signs inform you about how the covariate is related to moving up or down the scale (in logit form).
- ▶ We would want to compute odds ratios or probabilities.
- ▶ Either is simple to do and identical to what we've previously seen.
- ▶ Main complication with probabilities is that there are four of them.
- ▶ Consider Latino coefficient of -1.88 .
- ▶ Log-odds of responding in higher versus lower categories is -1.88 .

Ordinal Models: Ordinal Logit

- ▶ Odds ratio: $e^{-1.88} = .15$
- ▶ Odds less than 1 means that odds of responding in higher vs. lower categories is .15 that of non-Latino respondents.
- ▶ We can flip the interpretation. If $e^{-1.88}$ is odds of responding “up” the scale, then $(e^{-1.88})^{-1} = 1/e^{-1.88}$ gives the odds of responding in *lower* vs. *higher* categories.
- ▶ These odds are 6.52. This is easier to interpret: a Latino is 6.52 times more likely to respond in category $j - 1$ vs. category j .
- ▶ Probabilities? Simple.
- ▶ Let the covariate profile be: Democratic Latino respondent with mean income and mean education.
- ▶ This is covariate profile: PID: 1, Latino: 1, Income: 3, Education: 5

```
> z<-coef[1]*1 + coef[2]*5 + coef[3]*3 + coef[4]*1
>
> p1<- 1/(1+exp(z--1.3725485))
> p2<- 1/(1+exp(z-.2447428))-1/(1+exp(z--1.3725485))
> p3<- 1/(1+exp(z-1.5394910))-1/(1+exp(z-.2447428))
> p4<- 1-1/(1+exp(z-1.5394910))
>
> p1
[1] 0.5882365
> p2
[1] 0.2898005
> p3
[1] 0.08530296
> p4
[1] 0.03666003
```

Ordinal Models: Ordinal Logit

- ▶ These are the conditional probabilities.
- ▶ I chose the scenario.
- ▶ Should look a lot like binary logit, mechanically.
- ▶ These probabilities must sum to 1 (something has to happen!)
- ▶ Plots of probabilities are helpful. Odds ratio interpretation is also a useful thing to do here.
- ▶ What about probit?

```
> oprobit.mod<-polr(y ~ pid + education +
+ income + latino, method = c("probit"))
>
> summary(oprobit.mod)
```

Re-fitting to get Hessian

Call:

```
polr(formula = y ~ pid + education + income + latino, method = c("probit"))
```

Coefficients:

	Value	Std. Error	t value
pid	-0.34138575	0.07114005	-4.798784
education	0.01932492	0.02525182	0.765288
income	0.12393886	0.04420907	2.803472
latino	-1.03517869	0.13033518	-7.942435

Intercepts:

	Value	Std. Error	t value
1 2	-0.7489	0.1791	-4.1825
2 3	0.1839	0.1768	1.0401
3 4	0.9499	0.1784	5.3254

Residual Deviance: 1108.417

AIC: 1122.417

(6 observations deleted due to missingness)

```
>
```

Ordinal Models: Ordinal Probit

- ▶ Coefficients are of same sign and significance as logit.
- ▶ Interpretation is different.
- ▶ There is no odds ratio interpretation (same as binary probit).
- ▶ Consider probabilities for same scenario from before.

```
> z<-coef[1]*1 + coef[2]*5 + coef[3]*3 + coef[4]*1
>
> p1<probit(( -.7489-z), inverse=TRUE)
[1] FALSE
> p2<-probit((.1838660-z), inverse=TRUE)-probit((-.7489-z), inverse=TRUE)
> p3<-probit((.9499398-z), inverse=TRUE)-probit((.1838660-z), inverse=TRUE)
> p4<-1-probit((.9499398-z), inverse=TRUE)
>
>
> p1
[1] 0.5882365
> p2
[1] 0.2993276
> p3
[1] 0.1058389
> p4
[1] 0.03158003
>
>
```

Ordinal Models: Ordinal Probit

- ▶ Probabilities are essentially the same as in logit.
- ▶ The choice of models will generally be a matter of preference.
- ▶ I prefer logit because of odds ratio interpretation.
- ▶ And speaking of odds ratios . . .

Standard Practice with Ordinal Response Variables

- ▶ Assume Y is equal-interval scored and use normal-theory methods (OLS)
- ▶ Apply ordinal logits/probits but rarely evaluate prop. odds assumption
- ▶ Or do both, conduct “eyeball test”, and report OLS
- ▶ We can do better than this.
- ▶ Consider the following application from Jones and Westerland (unpublished)
- ▶ D.V. is attitudes on affirmative action (Likert-type scale).

Application

Table 1: Support for Affirmative Action
 OLS and Ordinal Logit Estimates

Variable	OLS Estimate	Proportional Odds Estimate
Symbolic Racism	1.375 (0.111)	2.562 (0.207)
Racial Prejudice	0.383 (0.193)	0.728 (0.356)
Education	0.259 (0.102)	0.510 (0.182)
Ideology	0.066 (0.027)	0.117 (0.047)
Female	0.025 (0.049)	0.075 (0.088)
Intercept	1.479 (0.127)	
α_1		0.652 (0.233)
α_2		1.888 (0.236)
α_3		3.268 (0.245)
ρ		1.744
log-likelihood:	1744	-2289.518

Non-proportionality

- ▶ Both ordinal logit/probit make the “parallel slopes” assumption.
- ▶ In logit setting, this is equivalent to the proportional odds property.
- ▶ It’s a “good” property insofar as if it holds, all we need to know about movement over the scale is $\hat{\beta}$.
- ▶ But what if it doesn’t hold?
- ▶ This may mean that one or more covariates has a differential effect over the range of the scale.
- ▶ Perhaps stronger effects above vs. below a midpoint.
- ▶ Proportionality tests are available.

Wald Test

- ▶ Wald test is a goodness-of-fit test.
- ▶ Useful for testing parameter restrictions.
- ▶ Consider the Wald χ^2 .
- ▶ Basic form:

$$\mathbf{Q}\boldsymbol{\beta} = \mathbf{r} \quad (8)$$

- ▶ \mathbf{Q} is a matrix of constants and \mathbf{r} is a vector of constants.
- ▶ $\boldsymbol{\beta}$ is the vector of parameters.

Wald Test

- ▶ Basic form:

$$W = [\mathbf{Q}\hat{\beta} - r]'[\mathbf{Q}\hat{\text{var}}(\hat{\beta})\mathbf{Q}']^{-1}[\mathbf{Q}\hat{\beta} - r] \quad (9)$$

- ▶ Suppose we have three parameters, β_0 , β_1 , and β_2 .
- ▶ Interested in $\beta_1 = \beta_2 = 0$.

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- ▶ In matrix form, this symbolizes what we're doing: are the coefficients jointly 0?
- ▶ Any other constraint could be imposed (single coefficient, multiple coefficients, etc.)

Wald Test

- ▶ For this test (without proof [see Long p. 91]),

$$W = \sum_{k=1}^2 z_{\hat{\beta}_k}^2$$

- ▶ z is the z ratio.
- ▶ The Wald test for proportionality is essentially evaluating cross-equation parameters for $J - 1$ logits.

Proportionality Tests

- ▶ Proportionality tests have been proposed:
Brant (*Biometrika*, 1990)
Peterson and Harrell (*Applied Statistics*, 1990)
- ▶ Basic concept, much simplified:
- ▶ Estimate $J - 1$ binary logits (1 if $y > m$; 0 if $y \leq m$)
- ▶ Extract estimated parameter vectors and covariance matrices from each model.
- ▶ Evaluate hypothesis that $\beta_{k1} = \beta_{k2} = \dots = \beta_{k,j-1}$
- ▶ Alex Mayer has coded this up in R; can do in Stata thanks to Scott Long.
- ▶ Essentially, a series of Wald tests over parameter restrictions.

Proportionality Tests

Table 2: Testing the Proportional Odds Assumption

Variable	Coefficient		
	$y > 1$	$y > 2$	$y > 3$
Symbolic Racism	1.822	2.458	3.037
Racial Prejudice	1.384	0.219	0.664
Education	0.209	0.496	0.651
Ideology	0.116	0.105	0.139
Female	-0.118	0.144	0.052
Constant	-0.279	-1.651	-3.645

Variable	χ^2	$p > \chi^2$	df
All	29.33	0.001	10
Symbolic Racism	12.37	0.002	2
Racial Prejudice	8.15	0.017	2
Education	2.29	0.319	2
Ideology	0.33	0.849	2
Female	5.79	0.055	2

Implications?

- ▶ Basic property of ordinal logit does not hold.
- ▶ Therefore, the model is inappropriate (statistically speaking)
- ▶ Substantively, you may be missing out on a lot of stuff!
- ▶ Why? Parameters may vary over the scale scores.
- ▶ This is potentially very interesting.
- ▶ Yet it almost always goes untested.
- ▶ There are easy to implement models.

Models: Nonproportional Odds

- ▶ Models for Nonproportional odds (nonparallel regression)
- ▶ Proposed most fully by Peterson and Harrell (1990); see also Williams (2006)
- ▶ Though McCullagh (1980) and others proposed such models.
- ▶ Basic idea: “let” regression parameters be unconstrained over scale scores.

Some Models for Nonproportional Odds

- ▶ Generalized Model:

$$\log \left[\frac{\Pr(Y \leq y_j | \mathbf{x})}{\Pr(Y > y_j | \mathbf{x})} \right] = \alpha_j - \mathbf{x}'\beta_j, \quad j = 1, 2, \dots, j-1$$

- ▶ Partial Proportional Odds:

$$\log \left[\frac{\Pr(Y \leq y_j | \mathbf{x})}{\Pr(Y > y_j | \mathbf{x})} \right] = -\alpha_j - \mathbf{x}'\beta - \mathbf{t}'\gamma_j, \quad j = 1, 2, \dots, j-1$$

- ▶ Restricted Generalized Logit:

$$\log \left[\frac{\Pr(Y \leq y_j | \mathbf{x})}{\Pr(Y > y_j | \mathbf{x})} \right] = \alpha_j + \mathbf{x}'\beta + \mathbf{z}'\zeta_j, \quad j = 1, 2, \dots, j-1$$

Unconstrained Partial Proportional Odds

- ▶ Originators: Peterson and Harrell, 1990.
- ▶ More Details: Probabilities

$$\Pr(Y \geq j | \mathbf{x}) = \frac{1}{1 + \exp(-\alpha_j - \mathbf{x}'\beta - \mathbf{t}'\gamma_j)}, \quad j = 1, 2, \dots, j-1$$

- ▶ \mathbf{t} are the q covariates exhibiting nonproportionality with associated parameter vector γ
 \mathbf{x} are p covariates having proportionality with associated parameter vector β
- ▶ Under original model, γ gives the “increment associated with the j th cumulative logit.” (P&H, 208).
- ▶ If $\gamma = 0$, proportionality holds and proportional odds model obtained.

Estimation and Implementation

- ▶ Log-Likelihood:

$$\log L = \sum_{i=1}^n \sum_{j=0}^k d_{ij} \log \Pr(Y = j \mid \mathbf{x}_i) = \sum_{i=1}^n \sum_{j=0}^k d_{ij} P_{ij}$$

- ▶ Has the same form as ordinal logit with difference that cumulative probabilities in UPP have nonproportional factors.
- ▶ Implementation
 1. Evaluate proportional odds assumption (global or covariate specific tests)
 2. Constrain covariates w/proportional odds to be equal over logits.
 3. Maximize log-likelihood.

Application: UPP

**Table 3: Support for Affirmative Action
Partial Proportional Odds (UPP Model)**

Variable	Coefficient	Standard Error
Symbolic Racism	1.788	(0.285)
γ_2	0.800	(0.257)
γ_3	1.345	(0.358)
Racial Prejudice	1.563	(0.495)
γ_2	-1.437	(0.440)
γ_3	-0.811	(0.586)
Female	-0.098	(0.125)
γ_2	0.277	(0.111)
γ_3	0.147	(0.145)
Education	0.479	(0.182)
Ideology	0.122	(0.047)
α_1	.273	(.22)
α_2	-1.635	(.203)
α_3	-3.263	(.236)
n	1744	
log-likelihood	-2274.293	

Application: Restricted Generalized Ordinal

Table 3a: Support for Affirmative Action
Partial Proportional Odds (Generalized Ordinal Logit)

Variable	Coefficient		
	C_1	C_2	C_3
Symbolic Racism	1.788 (0.285)	2.588 (0.245)	3.133 (0.280)
Racial Prejudice	1.563 (0.495)	0.127 (0.421)	0.752 (0.462)
Education	0.478 (0.182)	0.478 (0.182)	0.478 (0.182)
Ideology	0.122 (0.047)	0.122 (0.047)	0.122 (0.047)
Female	-0.098 (0.125)	0.180 (0.101)	0.049 (0.113)
Constant	-0.509 (0.283)	-1.698 (0.261)	-3.639 (0.299)
n	1744		
log-likelihood	-2274.293		

Summing Up Ordinality

- ▶ The OLS “strategy” is really bad in this context.
- ▶ Even the putatively correct model (ordinal logit/probit) may not hold.
- ▶ It must surely be true that if the logit model doesn't fit, OLS is even more awkward.
- ▶ Further, potentially interesting (substantively, theoretically, etc.) information is overlooked.
- ▶ Given prevalence of ordinal scales in research, these models would seem useful.

What about unordered response variables?

- ▶ Nominal categories without any natural ordering to them.
- ▶ Makes no sense to think about latency spanning the range of y .
- ▶ The “baseline” category or referent category is utterly arbitrary.
- ▶ Career decisions.
- ▶ Models for unordered outcomes are common.
- ▶ The most common is multinomial logit.

Multinomial Logit Logit

- ▶ The model is given by $J - 1$ nonredundant logits:

$$\log \left[\frac{\Pr(Y = y_j | \mathbf{x})}{\Pr(Y = 1 | \mathbf{x})} \right] = \alpha_j + \mathbf{x}'\beta_j, \quad j = 1, 2, \dots, j - 1$$

- ▶ As probabilities:

$$P_{ij} = \frac{\exp(\mathbf{x}'_i\beta_j)}{1 + \sum_{j=2}^J \exp(\mathbf{x}'_i\beta_j)}$$

- ▶ Likelihood:

$$\log L = \sum_{i=1}^n \sum_{j=1}^J d_{ij} \log P_{ij}$$

MNL

- ▶ Straightforward generalization of binary logit.
- ▶ Baseline category is utterly arbitrary.
- ▶ Many parameters: $J - 1 \times (K + 1)$
- ▶ Easy to impose constraints, if necessary.

MNL

- ▶ Let's see an application using an ordinal variable.
- ▶ This is ok to do.
- ▶ Item is on the effect immigrants will have on taking away jobs.
 - 1=extremely likely
 - 2=very likely
 - 3=somewhat likely
 - 4=not very likely
- ▶ 5 independent variables.
- ▶ Many parameters: $J - 1 \times (K + 1)$
- ▶ Easy to impose constraints, if necessary.
- ▶ In R I'm using VGAM

```
> imod_MNL<-vglm(formula= ImmJobs ~ Traits.Hispanics +
IDSelfPlace + gdiff + Morals + PersonalRetro,
+ family=multinomial); summary(imod_MNL)
```

Call:

```
vglm(formula = ImmJobs ~ Traits.Hispanics + IDSelfPlace + gdiff +
Morals + PersonalRetro, family = multinomial)
```

Pearson Residuals:

	Min	1Q	Median	3Q	Max
log(mu[,1]/mu[,4])	-2.3372	-0.36022	-0.26508	-0.16844	3.4832
log(mu[,2]/mu[,4])	-2.3859	-0.43538	-0.32315	-0.15246	2.5376
log(mu[,3]/mu[,4])	-2.4399	-0.58041	-0.50494	1.06772	1.4258

Coefficients:

	Value	Std. Error	t value
(Intercept):1	-4.74929980	0.811718	-5.8509217
(Intercept):2	-3.58840085	0.745811	-4.8114057
(Intercept):3	-1.35328556	0.639192	-2.1171814
Traits.Hispanics:1	3.43740323	1.062604	3.2348857
Traits.Hispanics:2	3.02004483	0.991489	3.0459685
Traits.Hispanics:3	1.92820198	0.875752	2.2017673
IDSelfPlace:1	-0.54782027	0.360729	-1.5186478
IDSelfPlace:2	-0.05799430	0.338796	-0.1711776
IDSelfPlace:3	-0.02699044	0.304273	-0.0887048
gdiff:1	0.03888022	0.106332	0.3656499
gdiff:2	0.02848386	0.099745	0.2855673
gdiff:3	0.00057201	0.087389	0.0065456
Morals:1	4.08198880	0.888800	4.5926961
Morals:2	3.26978148	0.831379	3.9329605
Morals:3	1.75242820	0.742368	2.3605932
PersonalRetro:1	1.21720657	0.475106	2.5619698
PersonalRetro:2	0.89297254	0.450565	1.9818959
PersonalRetro:3	0.71437247	0.409490	1.7445440

Number of linear predictors: 3

Names of linear predictors:

$\log(\mu_{1,1}/\mu_{1,4})$, $\log(\mu_{1,2}/\mu_{1,4})$, $\log(\mu_{1,3}/\mu_{1,4})$

Dispersion Parameter for multinomial family: 1

Residual Deviance: 1963.678 on 2304 degrees of freedom

Log-likelihood: -981.839 on 2304 degrees of freedom

Number of Iterations: 4

```
. mlogit ImmJobs Traits_Hispanics IDSelfPlace gdiff Morals PersonalRetro
```

Iteration 0: log likelihood = -1010.1141

Iteration 1: log likelihood = -982.54035

Iteration 2: log likelihood = -981.83989

Iteration 3: log likelihood = -981.83909

Iteration 4: log likelihood = -981.83909

Multinomial logistic regression

Number of obs = 774

LR chi2(15) = 56.55

Prob > chi2 = 0.0000

Pseudo R2 = 0.0280

Log likelihood = -981.83909

	ImmJobs	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
1						
Traits_His~s		1.509201	.8620393	1.75	0.080	-.1803648 3.198767
IDSelfPlace		-.5208298	.2879748	-1.81	0.071	-1.08525 .0435905
gdiff		.0383082	.0845181	0.45	0.650	-.1273442 .2039606
Morals		2.329561	.7070628	3.29	0.001	.9437431 3.715378

PersonalRe~o		.5028341	.3666761	1.37	0.170	-.2158378	1.221506
_cons		-3.396014	.676074	-5.02	0.000	-4.721095	-2.070934

2							
Traits_His~s		1.091843	.7789876	1.40	0.161	-.4349449	2.618631
IDSelfPlace		-.0310039	.2614533	-0.12	0.906	-.5434429	.4814351
gdif		.0279119	.0762702	0.37	0.714	-.121575	.1773987
Morals		1.517353	.6390465	2.37	0.018	.2648452	2.769861
PersonalRe~o		.1786001	.3363751	0.53	0.595	-.480683	.8378831
_cons		-2.235115	.6026392	-3.71	0.000	-3.416266	-1.053964

4							
Traits_His~s		-1.928202	.8758145	-2.20	0.028	-3.644767	-.2116373
IDSelfPlace		.0269905	.3042955	0.09	0.929	-.5694177	.6233987
gdif		-.000572	.0873925	-0.01	0.995	-.1718582	.1707141
Morals		-1.752428	.7424339	-2.36	0.018	-3.207572	-.2972848
PersonalRe~o		-.7143726	.4095318	-1.74	0.081	-1.51704	.0882949
_cons		1.353286	.6392393	2.12	0.034	.1003998	2.606172

(ImmJobs==3 is the base outcome)

MNL

- ▶ No difference ...
- ▶ except baseline category is chosen differently here.
- ▶ BC is category 3.
- ▶ Implications? None really.
- ▶ You can force Stata to change BC (base(4)) at end of model statement would do this.
- ▶ To see equivalence, note:

```
. display [1]_b[Traits_]-[4]_b[Traits_]
3.4374034
```

```
. display [2]_b[Traits_]-[4]_b[Traits_]
3.020045
```

```
. display -[4]_b[Traits_]
1.9282022
```

MNL

- ▶ These are differences in log-odds. In R the contrast is with category 4; in Stata, it's category 3.
- ▶ If you want to know log-odds of 1 vs. 4 (R), simply subtract log-odds of (4 vs. 3) from (1 vs. 3).

Issues

- ▶ Some issues with MNL
 - * Ordinality not preserved
 - * Contrasts of more natural interest not *directly* modeled.
 - * Though in principle, nothing is particularly wrong with MNL
- ▶ Suppose, however, we could reparameterize the MNL?
- ▶ Under MNL, the probabilities are given by:

$$P_{ij} = \frac{\exp(\mathbf{x}'_i\beta)}{1 + \sum_{j=2}^J \exp(\mathbf{x}'_j\beta)}$$

- ▶ Implies probability is assessed to baseline category.

An Alternative Parameterization

- ▶ Imagine a 4-point scale
- ▶ BCL (using category “1” as baseline):
Probability Contrasts: 4 vs. 1, 3 vs. 1, 2 vs. 1
- ▶ As an alternative, consider this:
Probability Contrasts: 4 vs. 3, 3 vs. 2, 2 vs. 1
- ▶ Here, we’re contrasting probabilities (or odds) with *adjacent categories*.
- ▶ This is the basic idea behind the adjacent category logit model.

Adjacent Category Logit Model

- ▶ The ideas of the previous slide can be summarized as:

$$\log \left(\frac{p_{j+1}}{p_j} \right) = \beta_j + \beta_j x_i.$$

- ▶ Coefficients are indexed by j implying a multinomial model.
- ▶ Probabilities are derived in terms of adjacent categories:

$$P_{j+1} \text{ vs. } j = \frac{\exp(Z_c)}{1 + \exp(Z_c)}$$

where Z_c corresponds to the linear predictor from the adjacent logit model.

- ▶ Possible to estimate single parameter vector (Agresti 1996)

```
> #Produces Adjacent Category Logit Estimates
>
> imod_ACL<-vglm(formula= ImmJobs ~ Traits.Hispanics + IDSelfPlace +
gdiff + Morals + PersonalRetro,
+ family=acat); summary(imod_ACL)
```

Call:

```
vglm(formula = ImmJobs ~ Traits.Hispanics + IDSelfPlace + gdiff +
Morals + PersonalRetro, family = acat)
```

Pearson Residuals:

	Min	1Q	Median	3Q	Max
log(P[Y=2]/P[Y=1])	-3.9584	0.10607	0.19399	0.445039	1.5449
log(P[Y=3]/P[Y=2])	-3.2526	-0.94751	0.39500	0.853176	1.7959
log(P[Y=4]/P[Y=3])	-1.2808	-0.59704	-0.22232	-0.084046	5.9901

Coefficients:

	Value	Std. Error	t value
(Intercept):1	0.990994	0.74849	1.32398
(Intercept):2	2.037344	0.60596	3.36217
(Intercept):3	1.304420	0.64866	2.01096
Traits.Hispanics:1	0.113170	0.95325	0.11872
Traits.Hispanics:2	-0.657034	0.79908	-0.82223
Traits.Hispanics:3	-1.914010	0.89626	-2.13556
IDSelfPlace:1	0.454362	0.31507	1.44208
IDSelfPlace:2	0.027242	0.26251	0.10378
IDSelfPlace:3	0.035982	0.30751	0.11701
gdiff:1	-1.276287	0.56082	-2.27577
gdiff:2	-1.373530	0.51677	-2.65789
gdiff:3	-0.575577	0.61926	-0.92946
Morals:1	-0.751327	0.78043	-0.96271
Morals:2	-1.442791	0.64219	-2.24668
Morals:3	-1.681574	0.74526	-2.25636
PersonalRetro:1	-0.320741	0.40503	-0.79190
PersonalRetro:2	-0.144079	0.33870	-0.42539

PersonalRetro:3 -0.685212 0.40960 -1.67288

Number of linear predictors: 3

Names of linear predictors:

$\log(P[Y=2]/P[Y=1])$, $\log(P[Y=3]/P[Y=2])$, $\log(P[Y=4]/P[Y=3])$

Dispersion Parameter for acat family: 1

Residual Deviance: 1931.95 on 2304 degrees of freedom

Log-likelihood: -965.975 on 2304 degrees of freedom

Number of Iterations: 4

Adjacent Category Logit Model

- ▶ Because model is reparameterized BCL, all fit statistics will be identical.
- ▶ There are no statistical grounds upon which to adjudicate one over the other.
- ▶ Illustration:
Log-odds for morality scale

$$\begin{aligned} C_2 \text{ vs. } C_1 &= -.751 \\ C_3 \text{ vs. } C_2 &= -1.443 \\ C_4 \text{ vs. } C_3 &= -1.682 \end{aligned}$$

Odds are:

$$\begin{aligned} C_2 \text{ vs. } C_1 &= \exp(-.751) = .471 \\ C_3 \text{ vs. } C_2 &= \exp(-1.443) = .236 \\ C_4 \text{ vs. } C_3 &= \exp(-1.682) = .186 \end{aligned}$$

- ▶ The contrasts are between adjacent categories.
- ▶ Interpretation? "We see that moral-traditionalists are less likely to respond in the higher categories as versus the lower categories. Further, this variable seems to not distinguish well categories 2 vs. 1. Interestingly, both categories represent the view that immigrants will *likely* (extremely or very) take away jobs."