Order Matters(?):: Alternatives to Conventional Practices for Ordinal Categorical Variables

Brad Jones

1Department of Political Science
University of California, Davis

February 9, 2007
Saints and Sinners

- All applied methods talks are really “saints” vs. “sinners.”
Saints and Sinners

- All applied methods talks are really “saints” vs. “sinners”.
- ...and today, I’m the preacher.
All applied methods talks are really “saints” vs. “sinners”.

... and today, I’m the preacher.

And I want to show you some paths to salvation.
The Sin

- Of course, there is variation on the severity of the sin.
The Sin

- Of course, there is variation on the severity of the sin.
- R vs. Stata
Of course, there is variation on the severity of the sin.

R vs. Stata

Bayesians vs. Frequentists
The Sin

- Of course, there is variation on the severity of the sin.
- R vs. Stata
- Bayesians vs. Frequentists
- Budweiser vs. Napa Valley Zinfandel
The Sin

- Of course, there is variation on the severity of the sin.
- R vs. Stata
- Bayesians vs. Frequentists
- Budweiser vs. Napa Valley Zinfandel
- And some of us preachers are a little more forgiving than others.
The sin I speak of today, many of you have probably committed.
The sin I speak of today, many of you have probably committed.

For it deals with ordinal dependent variables ...
The sin I speak of today, many of you have probably committed.

For it deals with ordinal dependent variables . . .

which are commonplace in applied social science work.
The sin I speak of today, many of you have probably committed.

For it deals with ordinal dependent variables . . .

which are commonplace in applied social science work.

Manifestations of the sin:
“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)

“Regression is remarkably robust in the face of violations of assumptions.” (unnamed former colleague).

“Since our two dependent variables have just four and five possible values, respectively, one option would be to present probit or logit results rather than ordinary least squares results presented in Table 2. As is well-known, however, the findings produced by these more complicated models are usually only trivially different from those obtained with OLS.” (Hibbing and Theiss-Morse, 1998 AJPS)

No justification given (Oliver and Mendelberg, 2000 AJPS)
The Origin of the Sin

- $y$ is understood to be limited (3-7 point scales, for example)
- Interest centers on $E(y_i) = f(x_i)$
- Equal-interval scoring on $y$ assumed.
- Least-squares solution is easy to get and to interpret.
- “a unit change in $x$ results in a $\beta$ change in $E(Y)$”
- Nobody gets hurt.
Committing sin in R:

```r
> modOLS<-lm(AAsupport ~ symracism + racialpred + Education + Ideology + female); summary(modOLS)

Call:
  lm(formula = AAsupport ~ symracism + racialpred + Education + Ideology + female)

Residuals:
       Min       1Q     Median       3Q       Max
-2.42160 -0.79044  0.13532  0.81650  2.31078

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.671030   0.099680  16.763  < 2e-16 ***
symracism     1.375278   0.110556  12.440  < 2e-16 ***
racialpred    0.191689   0.096593   1.985   0.0474 *
Education     0.259439   0.101971   2.544   0.0110 *
Ideology      0.066417   0.026633   2.494   0.0127 *
female        0.024651   0.049389   0.499   0.6178

---
Signif. codes:  *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 1.02 on 1738 degrees of freedom
(44 observations deleted due to missingness)
Multiple R-Squared: 0.1111, Adjusted R-squared: 0.1085
F-statistic: 43.45 on 5 and 1738 DF, p-value: < 2.2e-16
```
That apple tastes *soooo* good

▶ and yet you’ve sinned,
That apple tastes *soooo* good

- and yet you’ve sinned,
- and you know it.
That apple tastes *soooo* good

- and yet you’ve sinned,
- and you know it.
- “Because the dependent variables are categorical, OLS regression is technically inappropriate . . .”
Equal-interval scoring assumption is usually unrealistic.
Equal-interval scoring assumption is usually unrealistic.

Therefore, “a unit change in $x$ results in a $\beta$ change in $E(Y)$” may feel good to say...
Equal-interval scoring assumption is usually unrealistic.

Therefore, “a unit change in $x$ results in a $\beta$ change in $E(Y)$” may feel good to say...

But it may just be plain wrong.
Equal-interval scoring assumption is usually unrealistic.

Therefore, “a unit change in $x$ results in a $\beta$ change in $E(Y)$” may feel good to say...

But it may just be plain wrong.

Ordinality, conditional on covariates, may not exist.
Equal-interval scoring assumption is usually unrealistic.

Therefore, “a unit change in $x$ results in a $\beta$ change in $E(Y)$” may feel good to say...

But it may just be plain wrong.

Ordinality, conditional on covariates, may not exist.

Distinguishability between (among) scale categories may not occur among respondents.
Redemption?

- Assume $Y$ is a discretized measure of $Y^*$
- Postulate $Y_i^* = \alpha + \mathbf{x}'\beta + \epsilon_i$
- Cut Point Rule:
  
  \[
  \begin{align*}
    Y = 1 & \equiv Y^* \leq \alpha_1 \\
    Y = 2 & \equiv \alpha_1 < Y^* \leq \alpha_2 \\
    Y = 3 & \equiv \alpha_2 < Y^* \leq \alpha_3 \\
    Y = 4 & \equiv Y^* > \alpha_3
  \end{align*}
  \]
- Specify C.D.F. for $\epsilon$ (logistic, standard normal, cloglog)
- With logistic, proportional odds is obtained.
Redemption! The Proportional Odds Model

- Gives rise to:

\[
\Pr(Y \leq y_j \mid x) = \frac{\exp(\alpha_j - x' \beta)}{1 + \exp(\alpha_j - x' \beta)}
\]

- Linear Model for log-odds:

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

- Proportional Odds Property:

\[
\frac{\exp(x_1 \beta)}{\exp(x_2 \beta)} = \exp\{(x_1 - x_2)' \beta\}
\]
Repenting sin in R:

```r
> y<-factor(Aout$AAsupport)
> mod2<-polr(y~ symracism + racialpred + Education + Ideology + female); summary(mod2)
```

Re-fitting to get Hessian

Call:
polr(formula = y ~ symracism + racialpred + Education + Ideology + female)

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>symracism</td>
<td>2.5616408</td>
<td>0.20739086</td>
<td>12.3517542</td>
</tr>
<tr>
<td>racialpred</td>
<td>0.3638638</td>
<td>0.17798010</td>
<td>2.0444075</td>
</tr>
<tr>
<td>Education</td>
<td>0.5098697</td>
<td>0.18147497</td>
<td>2.8095871</td>
</tr>
<tr>
<td>Ideology</td>
<td>0.1172480</td>
<td>0.04717130</td>
<td>2.4855796</td>
</tr>
<tr>
<td>female</td>
<td>0.0753252</td>
<td>0.08798612</td>
<td>0.8561042</td>
</tr>
</tbody>
</table>

Intercepts:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>0.1807</td>
<td>1.5924</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.1827</td>
<td>8.3426</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.1932</td>
<td>15.0305</td>
</tr>
</tbody>
</table>

Residual Deviance: 4579.036
AIC: 4595.036
(44 observations deleted due to missingness)
Table 1: Support for Affirmative Action
OLS and Proportional Odds Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS</th>
<th>Proportional Odds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Standard Error)</td>
<td>(Standard Error)</td>
</tr>
<tr>
<td>Symbolic Racism</td>
<td>1.375</td>
<td>2.562</td>
</tr>
<tr>
<td></td>
<td>(0.111)</td>
<td>(0.207)</td>
</tr>
<tr>
<td>Racial Prejudice</td>
<td>0.383</td>
<td>0.728</td>
</tr>
<tr>
<td></td>
<td>(0.193)</td>
<td>(0.356)</td>
</tr>
<tr>
<td>Education</td>
<td>0.259</td>
<td>0.510</td>
</tr>
<tr>
<td></td>
<td>(0.102)</td>
<td>(0.182)</td>
</tr>
<tr>
<td>Ideology</td>
<td>0.066</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>Female</td>
<td>0.025</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.088)</td>
</tr>
<tr>
<td>Intercept</td>
<td>1.479</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.127)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.288</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.181)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>1.524</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.183)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>2.904</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.193)</td>
<td></td>
</tr>
<tr>
<td>$n$ log-likelihood:</td>
<td>1744</td>
<td>-2289.518</td>
</tr>
</tbody>
</table>

Order Matters(?)
Interocular Test (weak form): OLS $\approx$ P. Odds
Resolution, and Guilt Absolved

- Interocular Test (weak form): OLS \approx P. Odds
- Interocular Test (strong form): OLS > P. Odds
Resolution, and Guilt Absolved

- Interocular Test (weak form): OLS $\approx$ P. Odds
- Interocular Test (strong form): OLS $> P$. Odds
- And so it begins: proportional odds persecution.
Interocular Test (weak form): \( \text{OLS} \approx \text{P. Odds} \)

Interocular Test (strong form): \( \text{OLS} \succ \text{P. Odds} \)

And so it begins: proportional odds persecution.

“Because the ordinal logit results did not differ substantially from the OLS results, I present only the more easily interpretable OLS estimates.”

“. . . the more conventional OLS models are reported here.”

“logit coefficients lack intuitive interpretation.”

“OLS is chosen for simplicity of interpretation.”

“OLS is more accessible than logit”

“logit coefficients are not naturally interpretable”
Astray, the prodigal modeler publishes

- Deceived by the False Prophet, its “ease of interpretability.”
Astray, the prodigal modeler publishes

- Deceived by the False Prophet, its “ease of interpretability.”
- and yet, “similarity” may be illusory.
Astray, the prodigal modeler publishes

- Deceived by the False Prophet, its “ease of interpretability.”
- and yet, “similarity” may be illusory.
- Proportional odds (parallel regression) is an assumption
Deceived by the False Prophet, its “ease of interpretability.”

and yet, “similarity” may be illusory.

Proportional odds (parallel regression) is an assumption.

It may not hold. And if it doesn’t . . .
Astray, the prodigal modeler publishes

- Deceived by the False Prophet, its “ease of interpretability.”
- and yet, “similarity” may be illusory.
- Proportional odds (parallel regression) is an assumption
- It may not hold. And if it doesn’t . . .
- Any comparison between OLS and P. Odds is utterly meaningless.
Can I get a Witness?

- 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 $\beta_{k1} = \beta_{k2} = \ldots = \beta_{k,j-1}$
- Working on coding this up in R; can do in Stata thanks to Scott Long
The Sin, Exposed!

### Table 2: Testing the Proportional Odds Assumption

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Coefficient</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y &gt; 1$</td>
<td>1.822</td>
<td>2.458</td>
<td>3.037</td>
</tr>
<tr>
<td>$y &gt; 2$</td>
<td>1.384</td>
<td>0.219</td>
<td>0.664</td>
</tr>
<tr>
<td>$y &gt; 3$</td>
<td>0.209</td>
<td>0.496</td>
<td>0.651</td>
</tr>
<tr>
<td>Symbolic Racism</td>
<td>0.116</td>
<td>0.105</td>
<td>0.139</td>
</tr>
<tr>
<td>Racial Prejudice</td>
<td>-0.118</td>
<td>0.144</td>
<td>0.052</td>
</tr>
<tr>
<td>Education</td>
<td>-0.279</td>
<td>-1.651</td>
<td>-3.645</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\chi^2$</th>
<th>$p &gt; \chi^2$</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>29.33</td>
<td>0.001</td>
<td>10</td>
</tr>
<tr>
<td>Symbolic Racism</td>
<td>12.37</td>
<td>0.002</td>
<td>2</td>
</tr>
<tr>
<td>Racial Prejudice</td>
<td>8.15</td>
<td>0.017</td>
<td>2</td>
</tr>
<tr>
<td>Education</td>
<td>2.29</td>
<td>0.319</td>
<td>2</td>
</tr>
<tr>
<td>Ideology</td>
<td>0.33</td>
<td>0.849</td>
<td>2</td>
</tr>
<tr>
<td>Female</td>
<td>5.79</td>
<td>0.055</td>
<td>2</td>
</tr>
</tbody>
</table>

B.S. Jones

Order Matters(?)
And yet in despair, hope glimmers

- Models for Nonproportional odds (nonparallel regression)
- Proposed most fully by Peterson and Harrell (1990)
- Though McCullagh (1980) and others proposed such models.
- Basic idea: “let” regression parameters be unconstrained over scale scores.
Be Free! Some Models for Nonproportional Odds

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

- Partial Proportional Odds:
  \[
  \log \left[ \frac{\Pr(Y \leq y_j | x)}{\Pr(Y > y_j | x)} \right] = -\alpha_j - x' \beta - t' \gamma_j, \quad j = 1, 2, \ldots j - 1
  \]

- Restricted Generalized Logit:
  \[
  \log \left[ \frac{\Pr(Y \leq y_j | x)}{\Pr(Y > y_j | x)} \right] = \alpha_j + x' \beta + z' \zeta_j, \quad j = 1, 2, \ldots j - 1
  \]
Unconstrained Partial Proportional Odds

- More Details: Probabilities

\[
\Pr(Y \geq j \mid x) = \frac{1}{1 + \exp(-\alpha_j - x'\beta - t'\gamma_j)}, \quad j = 1, 2, \ldots, j-1
\]

- \(t\) are the \(q\) covariates exhibiting nonproportionality with associated parameter vector \(\gamma\)
- \(x\) are \(p\) covariates having proportionality with associated parameter vector \(\beta\)
- Under original model, \(\gamma\) 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:**

\[
\log L = \sum_{i=1}^{n} \sum_{j=0}^{k} d_{ij} \log \Pr(Y = j \mid x_i) = \sum_{i=1}^{n} \sum_{j=0}^{k} d_{ij} P_{ij}
\]

- 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.
## Application: UPP

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbolic Racism</td>
<td>1.788</td>
<td>(0.285)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.800</td>
<td>(0.257)</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>1.345</td>
<td>(0.358)</td>
</tr>
<tr>
<td>Racial Prejudice</td>
<td>1.563</td>
<td>(0.495)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-1.437</td>
<td>(0.440)</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-0.811</td>
<td>(0.586)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.098</td>
<td>(0.125)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.277</td>
<td>(0.111)</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.147</td>
<td>(0.145)</td>
</tr>
<tr>
<td>Education</td>
<td>0.479</td>
<td>(0.182)</td>
</tr>
<tr>
<td>Ideology</td>
<td>0.122</td>
<td>(0.047)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>.273</td>
<td>(.22)</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-1.635</td>
<td>(.203)</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>-3.263</td>
<td>(.236)</td>
</tr>
<tr>
<td>$n$</td>
<td>1744</td>
<td></td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-2274.293</td>
<td></td>
</tr>
</tbody>
</table>
Table 3a: Support for Affirmative Action
Partial Proportional Odds (Generalized Ordinal Logit)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbolic Racism</td>
<td>1.788</td>
<td>2.588</td>
<td>3.133</td>
</tr>
<tr>
<td></td>
<td>(0.285)</td>
<td>(0.245)</td>
<td>(0.280)</td>
</tr>
<tr>
<td>Racial Prejudice</td>
<td>1.563</td>
<td>0.127</td>
<td>0.752</td>
</tr>
<tr>
<td></td>
<td>(0.495)</td>
<td>(0.421)</td>
<td>(0.462)</td>
</tr>
<tr>
<td>Education</td>
<td>0.478</td>
<td>0.478</td>
<td>0.478</td>
</tr>
<tr>
<td></td>
<td>(0.182)</td>
<td>(0.182)</td>
<td>(0.182)</td>
</tr>
<tr>
<td>Ideology</td>
<td>0.122</td>
<td>0.122</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.047)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.098</td>
<td>0.180</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>(0.125)</td>
<td>(0.101)</td>
<td>(0.113)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.509</td>
<td>-1.698</td>
<td>-3.639</td>
</tr>
<tr>
<td></td>
<td>(0.283)</td>
<td>(0.261)</td>
<td>(0.299)</td>
</tr>
<tr>
<td>$n$</td>
<td>1744</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-2274.293</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A Sin Forgiven

- The OLS “strategy” is really bad in this context.
The OLS “strategy” is really bad in this context. Even the putatively correct model (ordinal logit/probit) may not hold.
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.
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.
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.
Yet No Good Deed Goes Unpunished

- Sometimes, ordinality is just an assumption.
Yet No Good Deed Goes Unpunished

- Sometimes, ordinality is just an assumption.
- Indeed, in nonproportional odds models, \( x \) may have "nonmonotonic" effects (i.e. odds or probabilities may not strictly increase or decrease wrt \( \Delta(x) \)).
Yet No Good Deed Goes Unpunished

- Sometimes, ordinality is just an assumption.
- Indeed, in nonproportional odds models, \( x \) may have “nonmonotonic” effects (i.e. odds or probabilities may not strictly increase or decrease wrt \( \Delta(x) \)).
- Of course the proportional odds/OLS strategy ensures this is the case (at least in the odds or \( E(y) \)).
Yet No Good Deed Goes Unpunished

- Sometimes, ordinality is just an assumption.
- Indeed, in nonproportional odds models, \( x \) may have “nonmonotonic” effects (i.e. odds or probabilities may not strictly increase or decrease wrt \( \Delta(x) \)).
- Of course the proportional odds/OLS strategy ensures this is the case (at least in the odds or \( E(y) \)).
- Maybe what you really need is an old friend to help you out.
What if a model for nominal categories were applied to a putatively ordinal scale?
In from the cold, a model returns

- What if a model for nominal categories were applied to a putatively ordinal scale?
- ... say, the multinomial logit model.
In from the cold, a model returns

▶ What if a model for nominal categories were applied to a putatively ordinal scale?
▶ ...say, the multinomial logit model.
▶ (Talk about logit persecution!)
What if a model for nominal categories were applied to a putatively ordinal scale?

...say, the multinomial logit model.

(Talk about logit persecution!)

In probabilities, the baseline category logit model:

\[ P_{ij} = \frac{\exp(x_i/\beta)}{1 + \sum_{j=2}^{J} \exp(x_i/\beta)} \]

Probability is assessed in reference to a baseline category.
BCL scorned: “my item is ordinal!”
A Model of Many Colors

- BCL scorned: “my item is ordinal!”
- Relax friend and imagine a 4-point scale
A Model of Many Colors

- BCL scorned: “my item is ordinal!”
- Relax friend and 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.
The ideas of the previous slide can be summarized as:

\[
\log \left( \frac{p_{j+1}}{p_j} \right) = \beta_j + \beta_j x_i.
\]
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, Yee and Hastie 2003)
```r
vglm(formula = ImmJobs ~ Traits.Hispanics + IDSelfPlace + gdiff + Morals + PersonalRetro, family = acat)

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept):1</td>
<td>0.990994</td>
<td>0.74849</td>
<td>1.32398</td>
</tr>
<tr>
<td>(Intercept):2</td>
<td>2.037344</td>
<td>0.60596</td>
<td>3.36217</td>
</tr>
<tr>
<td>(Intercept):3</td>
<td>1.304420</td>
<td>0.64866</td>
<td>2.01096</td>
</tr>
<tr>
<td>Traits.Hispanics:1</td>
<td>0.113170</td>
<td>0.95325</td>
<td>0.11872</td>
</tr>
<tr>
<td>Traits.Hispanics:2</td>
<td>-0.657034</td>
<td>0.79908</td>
<td>-0.82223</td>
</tr>
<tr>
<td>Traits.Hispanics:3</td>
<td>-1.914010</td>
<td>0.89626</td>
<td>-2.13556</td>
</tr>
<tr>
<td>IDSelfPlace:1</td>
<td>0.454362</td>
<td>0.31507</td>
<td>1.44208</td>
</tr>
<tr>
<td>IDSelfPlace:2</td>
<td>0.027242</td>
<td>0.26251</td>
<td>0.10378</td>
</tr>
<tr>
<td>IDSelfPlace:3</td>
<td>0.035982</td>
<td>0.30751</td>
<td>0.11701</td>
</tr>
<tr>
<td>gdiff:1</td>
<td>-1.276287</td>
<td>0.56082</td>
<td>-2.27577</td>
</tr>
<tr>
<td>gdiff:2</td>
<td>-1.373530</td>
<td>0.51677</td>
<td>-2.65789</td>
</tr>
<tr>
<td>gdiff:3</td>
<td>-0.575577</td>
<td>0.61926</td>
<td>-0.92946</td>
</tr>
<tr>
<td>Morals:1</td>
<td>-0.751327</td>
<td>0.78043</td>
<td>-0.96271</td>
</tr>
<tr>
<td>Morals:2</td>
<td>-1.442791</td>
<td>0.64219</td>
<td>-2.24668</td>
</tr>
<tr>
<td>Morals:3</td>
<td>-1.681574</td>
<td>0.74526</td>
<td>-2.25636</td>
</tr>
<tr>
<td>PersonalRetro:1</td>
<td>-0.320741</td>
<td>0.40503</td>
<td>-0.79190</td>
</tr>
<tr>
<td>PersonalRetro:2</td>
<td>-0.144079</td>
<td>0.33870</td>
<td>-0.42539</td>
</tr>
<tr>
<td>PersonalRetro:3</td>
<td>-0.685212</td>
<td>0.40960</td>
<td>-1.67288</td>
</tr>
</tbody>
</table>

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])

Log-likelihood: -965.975
```
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:
Odds for morality scale

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

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.
And yet, a hearty shock to the system . . .

- Multinomial models have multnomials! (Lots of parameters)
And yet, a hearty shock to the system . . .

- Multinomial models have multi nomials! (Lots of parameters)
- And we still can’t say much about ordinality.
And yet, a hearty shock to the system . . .

- Multinomial models have multi nominals! (Lots of parameters)
- And we still can’t say much about ordinality.
- Further, distinguishability between categories may not exist, given some set of covariates.
And yet, a hearty shock to the system . . .

- Multinomial models have multi nomials! (Lots of parameters)
- And we still can’t say much about ordinality.
- Further, distinguishability between categories may not exist, given some set of covariates.
- Respondents may not view category labels as meaningfully different (SA vs. A? So what!).
And yet, a hearty shock to the system . . .

- Multinomial models have multi nomials! (Lots of parameters)
- And we still can’t say much about ordinality.
- Further, distinguishability between categories may not exist, given some set of covariates.
- Respondents may not view category labels as meaningfully different (SA vs. A? So what!).
- The whole operation is falling apart.
Category labels on a scale may be regarded as “loose stereotypes.”
Category labels on a scale may be regarded as “loose stereotypes.”

Demarcations assumed by a researcher may not exist in the mind of the assessor/survey respondent.
Category labels on a scale may be regarded as “loose stereotypes.”

Demarcations assumed by a researcher may not exist in the mind of the assessor/survey respondent.

Yet non-orddinal models (like BCL) have many parameters (i.e. highly dimensional).
Category labels on a scale may be regarded as “loose stereotypes.”

Demarcations assumed by a researcher may not exist in the mind of the assessor/survey respondent.

Yet non-ordinal models (like BCL) have many parameters (i.e. highly dimensional).

Anderson derived the “stereotype” logit model (1984 Royal Stat. B.)
Category labels on a scale may be regarded as “loose stereotypes.”

Demarcations assumed by a researcher may not exist in the mind of the assessor/survey respondent.

Yet non-ordinal models (like BCL) have many parameters (i.e. highly dimensional).

Anderson derived the “stereoptype” logit model (1984 *Royal Stat. B.*).

It is really a canonical baseline category logit model.
Category labels on a scale may be regarded as “loose stereotypes.”

Demarcations assumed by a researcher may not exist in the mind of the assessor/survey respondent.

Yet non-ordinal models (like BCL) have many parameters (i.e. highly dimensional).

Anderson derived the “stereoptype” logit model (1984 Royal Stat. B.)

It is really a canonical baseline category logit model.

And hence, is a very general BCL.
Category labels on a scale may be regarded as “loose stereotypes.”

Demarcations assumed by a researcher may not exist in the mind of the assessor/survey respondent.

Yet non-ordinal models (like BCL) have many parameters (i.e. highly dimensional).

Anderson derived the “stereoptype” logit model (1984 *Royal Stat. B.*).

It is really a canonical baseline category logit model.

And hence, is a very general BCL.

Rarely has been applied; Greenland (1994) and esp. Yee and Hastie (2003) have revived it.
Stereotype Logit Model

Derived from the BCL:

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

As noted, ordinality is not built into this model.

Anderson postulated:

\[
\log \left[ \frac{\Pr(Y = y_j \mid x)}{\Pr(Y = 1 \mid x)} \right] = \alpha_j + x'\phi_j\beta, \quad j = 1, 2, \ldots j - 1
\]

Main differences?
Model is “one dimensional”: single parameter vector.
Parameters are weighted: \( \beta_j = \phi_j\beta \).
Stereotype Logit Model

- $\phi$ are scale parameters.
- For reduced rank models, they are weights on the $\beta_k$.
- Odds ratios are given by $\exp(\phi_j \beta_k)$ (note difference between this and BCL)
- Identification requires restrictions: $\phi_1 = 1, \phi_J = 0$
- Ordinality “exists” subject to constraint: $1 = \phi_1 > \phi_2 > \ldots \phi_J = 0$
- Distinguishability is now testable, given $\phi$.
- Model may be built up to incorporate multiple dimensions.
- Saturated model (full rank) is equivalent to BCL.
```
Call:
  rrvglm(formula = ImmJobs ~ Traits.Hispanics + IDSelfPlace +
          Morals + gdiff + PersonalRetro, family = multinomial, Rank = 1)
Coefficients:
     Value Std. Error t value
I(lv.mat):1  0.74862  0.094984  7.8816<---phi2
I(lv.mat):2  0.39212  0.084823  4.6228<---phi3 (note: phi1=1; phi4=0)
(Intercept):1  -4.28529  0.811086 -5.2834
(Intercept):2  -2.78453  0.692810  -4.0192
(Intercept):3   -0.60683  0.529454  -1.1461
Traits.Hispanics  2.37890  1.035743   2.2968
IDSelfPlace  -0.45178  0.337441  -1.3388
Morals         3.94622  0.850406    4.6404
gdiff          3.46905  0.682961    5.0794
PersonalRetro  0.96762  0.443165    2.1834
Number of linear predictors: 3
Names of linear predictors:
 log(mu[,1]/mu[,4]),  log(mu[,2]/mu[,4]),  log(mu[,3]/mu[,4])
 Log-likelihood:  -969.5284
```
Interpretation

- Note restrictions on $\phi$
- $\phi_1 = 1$, $\phi_4 = 0$
- Ordinality constraints seem to hold: $\phi_1 > \phi_2 > \phi_3 > \phi_4$.
- Given statistical significance of $\phi$, distinguishability condition seems to hold.
- $\phi$ conveys information about the relative difference in weights associated with the covariates.
- Covariate effects are largest for “not very likely” vs. “somewhat likely” categories. (4 vs. 3)
- The difference is about .39 on the log odds scale. Difference between 2 vs. 1 is about .25.
- Consider the morals scale:

  Log-odds for morality scale

  $C_1$ vs. $C_4 = 3.95$
  $C_2$ vs. $C_4 = 3.95 \times .749 = 2.96$
  $C_3$ vs. $C_4 = 3.95 \times .392 = 1.55$

  Odds are:

  $C_1$ vs. $C_4 = \exp(3.95) \approx 51$
  $C_2$ vs. $C_4 = \exp(2.96) \approx 19$
  $C_3$ vs. $C_4 = \exp(1.55) \approx 5$

- Clearly a close connection to the BCL
- BCL does not fit better than this model. We might prefer to report this model.
Stereotype model gives a test for ordinality and distinguishability

It also allows for implicit comparisons with higher dimensional models.

If it holds, it has the desirable feature of being reduced rank (previous slide: rank=1)

Ultimately, what should you do?

Try a variety of fits (Gelman and Hill 2007).

Take seriously the trinity of assumptions:
- Proportionality (parallelism)
- Ordinality
- Distinguishability

All of the models considered permit some leverage on these issues.
Stereotype model gives a test for ordinality and distinguishability.

It also allows for implicit comparisons with higher dimensional models.

If it holds, it has the desirable feature of being reduced rank (previous slide: rank=1).

Ultimately, what should you do?

Try a variety of fits (Gelman and Hill 2007).
... Revelation

- Stereotype model gives a test for ordinality and distinguishability.
- It also allows for implicit comparisons with higher dimensional models.
- If it holds, it has the desirable feature of being reduced rank (previous slide: rank=1).
- Ultimately, what should you do?
- Try a variety of fits (Gelman and Hill 2007).
- Take seriously the trinity of assumptions: Proportionality (parallelism), Ordinality, Distinguishability.
Stereotype model gives a test for ordinality and distinguishability.

It also allows for implicit comparisons with higher dimensional models.

If it holds, it has the desirable feature of being reduced rank (previous slide: rank=1).

Ultimately, what should you do?

Try a variety of fits (Gelman and Hill 2007).

Take seriously the trinity of assumptions: Proportionality (parallelism)
Ordinality
Distinguishability

All of the models considered permit some leverage on these issues.
The Neverending Quest

- R functionality needs to be improved (working on it)
- Bayesian framework ideal for the setting considered here (kinda thinkin’ about it)
- Lots of room for reanalysis (of the positive kind)