Ordinal Models

Brad Jones[1]

[1]Department of Political Science
University of California, Davis

January 24, 2007

Today: Ordinal and Multinomial Models

Topics of Course

- Alternatives to Proportional Odds Model

Topics of Course

- Alternatives to Proportional Odds Model
- Multinomial Models and Alternatives.

Unconstrained Partial Proportional Odds

- Originators: Peterson and Harrell, 1990.

Unconstrained Partial Proportional Odds

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

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

Unconstrained Partial Proportional Odds

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

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

- $\mathbf{t}$ are the q covariates exhibiting nonproportionality with associated parameter vector $\gamma(q \times 1)$
  $\mathbf{x}$ are p covariates having proportionality with associated parameter vector $\beta(p \times 1)$

Unconstrained Partial Proportional Odds

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

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

- $\mathbf{t}$ are the q covariates exhibiting nonproportionality with associated parameter vector $\gamma(q \times 1)$
  $\mathbf{x}$ are p covariates having proportionality with associated parameter vector $\beta(p \times 1)$
- Under original model, γ gives the "increment associated with the jth cumulative logit." (P&H, 208).

Unconstrained Partial Proportional Odds

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

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

- $\mathbf{t}$ are the q covariates exhibiting nonproportionality with associated parameter vector $\gamma(q \times 1)$
  $\mathbf{x}$ are p covariates having proportionality with associated parameter vector $\beta(p \times 1)$
- Under original model, γ gives the "increment associated with the jth cumulative logit." (P&H, 208).
- If $\gamma = 0$, proportionality holds and proportional odds model obtained.

Estimation and Implementation

- Log-Likelihood:

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

Estimation and Implementation

- Log-Likelihood:

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

- 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.

Implementation

- In R, Yee's `vgam` environment will give fully generalized ordinal logit.
  In Stata, `gologit2` will produce the model and test. (I'm working on implementation of model in R).

Implementation

- In R, Yee's vgam environment will give fully generalized ordinal logit.
  In Stata, gologit2 will produce the model and test. (I'm working on implementation of model in R).
- For Stata annotated do file, go to:
  http://psfaculty.ucdavis.edu/bsjjones/ordermatters.txt

Multinomial Models

- Sometimes ordinality will not hold (or will not be of primary interest)
- Different contrasts may be of interest
- Leads to consideration of alternative models

Baseline Category Logit

- By way of review, consider the baseline category logit:

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

Baseline Category Logit

- By way of review, consider the baseline category logit:

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

- As probabilities:

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

Baseline Category Logit

- By way of review, consider the baseline category logit:

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

- As probabilities:

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

- Likelihood:

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

Baseline Category Logit

- Straightforward generalization of binary logit.

Baseline Category Logit

- Straightforward generalization of binary logit.
- Baseline category is utterly arbitrary.

Baseline Category Logit

- Straightforward generalization of binary logit.
- Baseline category is utterly arbitrary.
- Many parameters: $J - 1 \times (K + 1)$

Baseline Category Logit

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

Implementation

An illustration of the BCL can be found at:

http://psfaculty.ucdavis.edu/bsjjones/multinomialmodels.pdf

Issues

- Some issues with BCL
  * Ordinality not preserved
  * Contrasts of more natural interest not *directly* modeled.
  * Though in principle, nothing is particularly wrong with BCL

Issues

- Some issues with BCL
  - * Ordinality not preserved
  - * Contrasts of more natural interest not *directly* modeled.
  - * Though in principle, nothing is particularly wrong with BCL
- Suppose, however, we could reparameterize the BCL?

Issues

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

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

Issues

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

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

- Implies probability is assessed to baseline category.

An Alternative Parameterization

- Imagine a 4-point scale

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

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

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.*

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:

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

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.

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:

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

where Z_c corresponds to the linear predictor from the adjacent 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:

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

  where Z_c corresponds to the linear predictor from the adjacent logit model.
- Possible to estimate single parameter vector (Agresti 1996)

Implementation

An illustration of the BCL can be found at:

http://psfaculty.ucdavis.edu/bsjjones/multinomialmodels.pdf

Considerations of Ordinality

- Ordinality with respect to covariates may not always hold.

Considerations of Ordinality

- Ordinality with respect to covariates may not always hold.
- Distinguishability between categories may not exist, given some set of covariates.

Considerations of Ordinality

- Ordinality with respect to covariates may not always hold.
- Distinguishability between categories may not exist, given some set of covariates.
- Respondents may not view category labels as meaningfully different.

Considerations of Ordinality

- Ordinality with respect to covariates may not always hold.
- Distinguishability between categories may not exist, given some set of covariates.
- Respondents may not view category labels as meaningfully different.
- This in turn may call into question assumptions about ordinality of the response variable.

Anderson's Stereotype Logit Model

- Category labels may be "loose stereotypes."

Anderson's Stereotype Logit Model

- Category labels may be "loose stereotypes."
- Demarcations assumed by a researcher may not exist in the mind of the assessor.

Anderson's Stereotype Logit Model

- Category labels may be "loose stereotypes."
- Demarcations assumed by a researcher may not exist in the mind of the assessor.
- Anderson derived the "stereoptype" logit model (1984 *Royal Stat. B.*)

Anderson's Stereotype Logit Model

- Category labels may be "loose stereotypes."
- Demarcations assumed by a researcher may not exist in the mind of the assessor.
- Anderson derived the "stereoptype" logit model (1984 *Royal Stat. B.*)
- It is really a canonical baseline category logit model.

Stereotype Logit Model

- Derived from the BCL:

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

Stereotype Logit Model

- Derived from the BCL:

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

- As we've noted, ordinality is not built into this model.

Stereotype Logit Model

- Derived from the BCL:

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

- As we've noted, ordinality is not built into this model.
- Anderson postulated:

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

Stereotype Logit Model

- Derived from the BCL:

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

- As we've noted, ordinality is not built into this model.
- Anderson postulated:

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

- Main differences?
  Model is "one dimensional": single parameter vector.
  Parameters are weighted: $\beta_j = \phi_j\beta$.

Stereotype Logit Model

- ϕ are scale parameters.

Stereotype Logit Model

- ϕ are scale parameters.
- They are weights on the parameters.

Stereotype Logit Model

- ϕ are scale parameters.
- They are weights on the parameters.
- Identification requires restrictions: $\phi_1 = 1, \phi_J = 0$

Stereotype Logit Model

- ϕ are scale parameters.
- They are weights on the parameters.
- Identification requires restrictions: $\phi_1 = 1, \phi_J = 0$
- Ordinality "exists" subject to constraint:
  $1 = \phi_1 > \phi_2 > \ldots \phi_J = 0$

Stereotype Logit Model

- ϕ are scale parameters.
- They are weights on the parameters.
- 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 ϕ.

Stereotype Logit Model

- ϕ are scale parameters.
- They are weights on the parameters.
- 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 ϕ.
- Model may be built up to incorporate multiple dimensions.

Stereotype Logit Model

- ϕ are scale parameters.
- They are weights on the parameters.
- 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 ϕ.
- Model may be built up to incorporate multiple dimensions.
- Saturated model is equivalent to BCL.

Implementation

An illustration of the stereotype logit can be found at:
http://psfaculty.ucdavis.edu/bsjjones/multinomialmodels.pdf
.