UNIVERSITY OF
SOUTH CAROLINA

IRT MODELING OF ORDINAL FORCED-CHOICE DATA

Alberto Maydeu Olivares

University of South Carolina

Columbia, SC
October 13, 2016

LATENT VARIABLE MODELS FOR CHOICE DATA

- We often model choice data using latent variable models

- Sometimes it is not obvious that it is choice data

If $5x + 32 = 4 - 2x$ what's the value of x?
A) -4 B) -3 C) 4 D) 7 E) 12

- Others, it is pretty obvious

Your parents insist that you select a major that you are not interested in.
A) You select the major without complaining
B) You try to persuade your parents that the major is not for you
C) You talk to your relatives so that they help you convince your parents not to select it

- If you ask individuals to choose one answer, how do you model the data?
  - Multinomial logistic regression where the predictor is a latent variable
    - Mathematical ability or assertiveness
  - The model was proposed by Bock (1972)

LATENT VARIABLE MODELS FOR CHOICE DATA

- Choice data is somewhat popular in personnel assessment because of concerns with faking

	Very true of me	Somewhat true of me	Somewhat untrue of me	Very untrue of me
A. I manage to relax easily	x			
B. I am careful over detail		x		
C. I enjoy working with others			x	
D. I set high personal standards			x	

	Ranking	Partial ranking	First choice
A. I manage to relax easily	1	1	1
B. I am careful over detail	2		
C. I enjoy working with others	4	4	
D. I set high personal standards	3		

- Loss of information + missing data problem

LATENT VARIABLE MODELS FOR CHOICE DATA

- Generalized linear models with latent variable predictors
  - Logistic or normal CDF link function
- Let's focus on the linear part
- First, how do we code the data?

Partial ranking				Binary Outcomes					
A	B	C	D	{A,B}	{A,C}	{A,D}	{B,C}	{B,D}	{C,D}
most		least		1	1	1	1	.	0

LINEAR PART

- $\mathbf{t}$ is a n vector of utilities; $\mathbf{y}^*$ is a $\tilde{n} = \frac{n(n-1)}{2}$ vector of latent difference responses
- For ranking data, we write the set of $\tilde{n}$ equations as $\mathbf{y}^* = \mathbf{A}\,\mathbf{t}$,

where $\mathbf{A}$ is a $\tilde{n} \times n$ design matrix. For n = 2, 3, 4,

$$\mathbf{A} = \begin{pmatrix} 1 & -1 \end{pmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix},$$

- For paired comparison data you need $\mathbf{y}^* = \mathbf{A}\,\mathbf{t} + \mathbf{e}$
- m common factors (latent traits) underlie the utilities $\mathbf{t}$

$$\mathbf{t} = \mu_t + \Lambda\eta + \varepsilon$$

- It is a second order factor analysis model for binary data

$$\mathbf{y}^* = \mathbf{A}\left(\mu_t + \Lambda\eta + \varepsilon\right)$$

FUNDAMENTAL PROBLEM WITH MODELS FOR CHOICE DATA

- An assertiveness item

  Your parents insist that you select a major that you are not interested in.
  A) You select the major without complaining
  B) You try to persuade your parents that the major is not for you
  C) You talk to your relatives to help you convince your parents not to select it

- Model: $\mathbf{y}^* = \mathbf{A}\left(\mu_t + \Lambda\eta + \varepsilon\right)$

$$\mathbf{A}\lambda = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} .1 \\ .8 \\ .6 \end{bmatrix} = \begin{bmatrix} -.7 \\ -.5 \\ .2 \end{bmatrix}$$

- When alternatives reflect only **one** latent variable resulting slopes need to be low
  - Unless one original slope is 0, or better yet negative!

- Unless very carefully constructed, choice data will yield very little information on the latent traits
  - Item parameter estimation will be difficult
  - Many items needed to accurately estimate the latent variables (individual scores)

FUNDAMENTAL PROBLEM WITH MODELS FOR CHOICE DATA

- The problem has a difficult solution for models for mathematics data
  - Alternatives reflect a single latent variable
  - Do we want to include alternatives with negative slopes? 0 slopes?

If $5x + 32 = 4 - 2x$ what's the value of x?

A) -4 B) -3 C) 4 D) 7 E) 12

- Problem has an easier solution in personality/attitude assessment
  - Combine alternatives that measure different latent variables

	Ranking	Partial ranking	First choice
A. I manage to relax easily	1	1	1
B. I am careful over detail	2		
C. I enjoy working with others	4	4	
D. I set high personal standards	3		

COMBINING ALTERNATIVES FOR DIFFERENT TRAITS

- 3 attributes, each measured by 3 items, presented in triplets
- Thurstonian factor model: $\mathbf{y}^* = \mathbf{A}(\mu_t + \Lambda\eta + \varepsilon)$, with $\mathrm{var}(\eta) = \mathbf{\Phi}, \mathrm{var}(\varepsilon) = \mathbf{\Psi}^2$

$$\mathbf{\Lambda} = \left(\begin{array}{ccc} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \\ \hline \lambda_4 & 0 & 0 \\ 0 & \lambda_5 & 0 \\ 0 & 0 & \lambda_6 \\ \hline \lambda_7 & 0 & 0 \\ 0 & \lambda_8 & 0 \\ 0 & 0 & \lambda_9 \end{array}\right) \quad \mathbf{A} = \left(\begin{array}{ccc|ccc|ccc} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{array}\right)$$

$$\mathbf{\Psi}^2 = diag\left(\psi_1^2 \quad \psi_2^2 \quad 1 \;\vert\; \psi_4^2 \quad \psi_5^2 \quad 1 \;\vert\; \psi_7^2 \quad \psi_8^2 \quad 1\right),\ \mu_t = (\mu_1, \cdots, \mu_8, 0),\ \mathbf{\Phi}$$

REMAINING PROBLEMS

- Still a lot of work to construct the questionnaires
- Data is binary (dummy variables), you need more items to reach the same precision of measurement than with graded ratings
- How to compete with graded ratings? Graded paired comparisons

A	Much better	Somewhat Better	Somewhat better	Much better	B
I manage to relax easily	○	●	○	○	I am careful over detail

MODELING ORDINAL PAIRED COMPARISONS

- Thurstonian factor model: $\mathbf{y}^* = \mathbf{A}\left(\mu_t + \Lambda\eta + \varepsilon\right) + \mathbf{e}$, with $\text{var}\left(\eta\right) = \mathbf{\Phi}, \text{var}\left(\varepsilon\right) = \mathbf{\Psi}, \text{var}\left(\mathbf{e}\right) = \mathbf{\Theta}$

- $E\left(\mathbf{y}^*\right) = \mathbf{A}\mu_t = \gamma \qquad \text{var}\left(\mathbf{y}^*\right) = \mathbf{A\Lambda\Phi\Lambda'A'} + \mathbf{A\Psi A'} + \mathbf{\Theta} = \breve{\mathbf{\Lambda}}\mathbf{\Phi}\breve{\mathbf{\Lambda}}' + \breve{\mathbf{\Psi}} + \mathbf{\Theta}$

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ \hline 0 & 0 & \lambda_3 \\ \lambda_4 & 0 & 0 \\ \hline 0 & \lambda_5 & 0 \\ 0 & 0 & \lambda_6 \\ \hline \vdots & \vdots & \vdots \end{pmatrix}, \quad \breve{\mathbf{\Lambda}} = \begin{pmatrix} \lambda_1 & -\lambda_2 & 0 \\ -\lambda_4 & 0 & \lambda_3 \\ 0 & \lambda_5 & -\lambda_6 \\ \lambda_7 & -\lambda_8 & 0 \\ -\lambda_{10} & 0 & \lambda_9 \\ 0 & \lambda_{11} & -\lambda_{12} \end{pmatrix} \breve{\mathbf{\Psi}} = \begin{pmatrix} \psi_1 + \psi_2 & & & \\ 0 & \psi_3 + \psi_4 & & \\ \vdots & & \ddots & \\ 0 & 0 & & \psi_{2n-1} + \psi_{2n} \end{pmatrix}$$

- Elements of $\mathbf{\Psi}$ are not identified, set them to ½.
- Graded paired comparisons provide half the information that graded ratings
  - o But items may be repeated in different pairs
  - o It is parsimonious to use $\mathbf{\Theta} = \omega\mathbf{I}$

STATISTICAL CONSIDERATIONS

- **Link function**: normal (greater flexibility)

- **Model**: $\mathbf{z}^* = \mathbf{D}\left(\mathbf{y}^* - \mu_{y^*}\right)$, with $\mathbf{D} = \left(\text{Diag}\left(\Sigma_{y^*}\right)\right)^{-\frac{1}{2}}$

  with $y_i = k_i$ if $\tau_{i_k} < z_i^* < \tau_{i_{k+1}}$, $k_i = 0, \ldots, K - 1$, where $\tau_{i_0} = -\infty, \tau_{i_K} = \infty$

  - Multivariate ordinal probit regression with latent variable predictors
    - With a very special constraints

- **Estimation**
  - Bayesian
  - ML (not feasible)
  - Limited information (polychorics, bivariate composite likelihood, GEE)

- **Scoring** Empirical Bayes modal (EBM/MAP)

STATISTICAL CONSIDERATIONS

- We use methods based on polychorics
  1) estimate each threshold separately by ML
  2) estimate each polychoric correlation separately by pseudo ML
  3) estimate model parameters by minimum distance estimation using the estimated thresholds and polychorics

- CAN estimates
- Asymptotically correct SEs and goodness of fit tests
  - How well the model reproduces the bivariate tables

- In practice, methods based on polychorics are as efficient as ML

- No estimation queries, just modelling queries
  - How to best combine items so that they provide maximum information
    - Reduce test length

- Perhaps we would like a better goodness of fit test….

UNIVERSITY OF SOUTH CAROLINA

APPLICATION

- Big Five questionnaire
  - 60 items measuring 5 psychological traits
    - 12 items per trait

- 20 blocks of 3 items
  - Items presented within the block using 3 graded paired comparisons

- *N* = 596

- After completing the questionnaire, individuals were presented with the same items in rating format

THANK YOU FOR YOUR ATTENTION