Advances in Models for Multivariate Nominal or Ordinal Variables with Latent Variables

IMPS
July 2010
Athens Georgia

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Models for Nominal & Ordinal Variables

- Example data set: Espelage et al.
- Review of existing approaches for analyzing such data.
- Log-multiplicative association models.
  - Underlying models that lead to LMA
  - Conditional Specification.
- The State of the Art:
  - Multiple correlated latent variables.
  - Restrictions on scale values for response options and location parameters.
  - Covariates.
  - Estimation.
- Areas for future work.
- Time permitting, a nominal example.

The Espelage, Holt & Henkel Data (2004)

Data from Espelage, D.L., Holt, M.K., & Henkel, R.R. (2004). Examination of peer-group contextual effects on aggression during early adolescence. *Child Development*, 74, 205–220.

The data consist of responses of students from a midwestern middle school to items of the Illinois Victimization Scale (Espelage & Holt, 2001).

Bully Sub-scale Items:
- I upset other students for the fun of it.
- I helped harass other students.
- I teased other students.

Fight Sub-scale Items:
- I got in a physical fight.
- I threatened to hurt or hit another student.
- I hit back when someone hits me first.

Response scale:

Never, 1 or 2 times, 3 or 4 times, 5 or 6 times, and 7 or more times.

The role Gender? Do girls tend to be more verbal bullies and boys more physical? *. . . The findings are mixed. . .*

Review of Major Existing Approaches

to Latent Variable Modeling of Discrete Response Data:

- Quantify the data and then use factor analysis (or SEM) for continuous data.
  - Need to know the order of the response options.
  - Does not allow for alternative scoring for different latent variables.
- Item response theory models for polytomous items
  - Multiple latent variables is a problem for standard estimation algorithms (i.e., numerical integration).
- Factor analysis of discrete data (Bartholomew, Steele, Moustaki & Galbraith, 2008)
  - Lack of available of software and flexibility of implementation.
  - Methods and programs for nominal data are sorely lacking and "...work on ordinal categorical variables is nearer the research frontier and is consequently more incomplete, and in some sense, more difficult than other methods." (p. 243)

Log-multiplicative Association Models

- Structured Poisson regression model with 2-way interactions.
- Generalization of Goodman's (1979, 1986) $RC(M)$ association model for two-way tables to mulit-way tables, i.e.,

$$\log(P(y_i = j_i, y_k = \ell_k)) = \lambda + \lambda^R_{ij_i} + \lambda^C_{k\ell_k} + \sum_m \phi_m \nu^R_{ij_im}\nu^C_{k\ell_km}$$

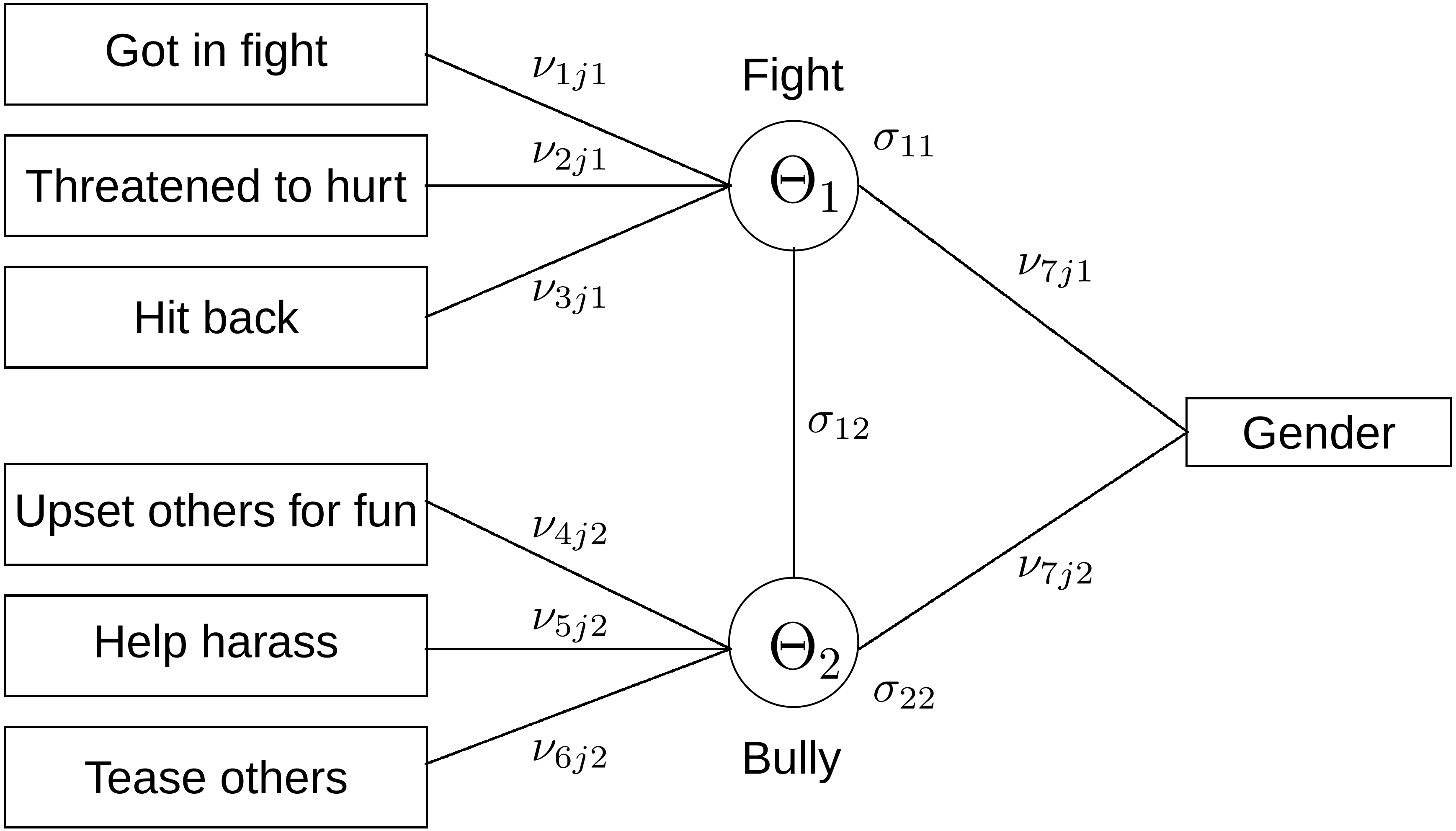
- When equally spaced scores are input for the ν's (and $M = 1$), then the model is known as a the uniform association model.
- Takane (1987): Ideal point discriminant analysis without a centroid restriction on the columns (criterion groups) is equivalent to the RC association model.
- Andersen (1995): Rasch model for polytomous items where an item's response options are the rows and the columns are (categorical) estimates of ability/latent trait.
- The general log-multiplicative association (LMA) model

$$\log(P(\boldsymbol{y})) = \lambda + \sum_i \lambda_{ij} + \sum_i \sum_{k>i} \sum_m \sum_{m' \geq m} \sigma_{mm'}\nu_{ijm}\nu_{k\ell m'}$$

Underlying Models that Imply a LMA Model

that I know of. . .

- Are implied by underlying multivariate normal distribution (Goodman, 1979; Becker, 1989; others).
- Can be derived from a distance based model (de Rooij & Heiser, 2005), and a generalization of Newton's Law of Gravity (de Rooij, 2008).
- For dichotomous items, derived via the Dutch Identity (Holland, 1993).
- A generalization of the Dutch Identity to Rasch models for polyotmous items, multidimensional traits and covariates. (Li, 2010).
- Can be derived from a formative latent variable model using statistical graphical models (Anderson & Böckenholt, 2000; Anderson & Vermunt, 2000; Anderson, 2002; Anderson & Tettagah, 2007).
- Can be derived from a reflective latent variable model using standard item response theory methodology (Anderson & Yu, 2007; Anderson, Li, & Vermunt, 2007; Anderson, Verkuilen & Peyton, in press).



$$\log(P(\boldsymbol{y})) = \lambda + \sum_{i=1}^{7} \lambda_{ij} + \sum_{i=1}^{7} \sum_{k>i} \sum_{m=1}^{2} \sum_{m' \geq m} \sigma_{mm'} \nu_{ijm} \nu_{k\ell m'}$$

Assumptions & Implications

- The response pattern $\boldsymbol{y}$ follows a multinomial distribution.
- Absence of a line connecting variables indicates conditional independence
- For all possible response patterns $\boldsymbol{y}$,

$$\boldsymbol{\Theta}|\mathbf{y} \sim MVN\left(\boldsymbol{\mu}_{\mathbf{y}}, \boldsymbol{\Sigma}\right) \ \ i.i.d.$$

where

$$\boldsymbol{\mu}_y = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1M} & \sigma_{2M} & \cdots & \sigma_{MM} \end{pmatrix} \begin{pmatrix} \sum_i \nu_{ij1} \\ \sum_i \nu_{ij2} \\ \vdots \\ \sum_i \nu_{ijM} \end{pmatrix}$$

Our example,

Fight: $\mu_1|\mathbf{y} = \sigma_{11}(\sum_i \nu_{ij1}) + \sigma_{12}(\sum_i \nu_{ij2})$

Bully: $\mu_2|\mathbf{y} = \sigma_{22}(\sum_i \nu_{ij2}) + \sigma_{12}(\sum_i \nu_{ij1})$

Conditional Approach

Consider the following conditional logistic regression model,

$$P(Y_i = j | y_{k\ell}, k \neq i, \boldsymbol{x}) = \frac{\exp(\lambda_{ij} + \sum_p \beta_{ijp} x_p + \sum_{k \neq i} \psi_{ij|k\ell})}{\sum_h \exp(\lambda_{ih} + \sum_p \beta_{ihp} x_p + \sum_{k \neq i} \psi_{ih|k\ell})}$$

where

- λ_{ij} is an intercept or location parameter.
- $\psi_{ij|k\ell}$ is the parameter for variable y_k when predicting variable y_i.

If we have this model for each item i ($i = 1, \ldots, I$) and $\psi_{ij|k\ell} = \psi_{k\ell|ij}$, then model for the joint distribution of all items is

$$\log(P(y_{1j}, \ldots, y_{Ij} | \boldsymbol{x})) = \lambda + \sum_i \lambda_{ij} + \sum_i \sum_p \beta_{ijp} x_p + \sum_i \sum_{k>i} \psi_{ij|k\ell}$$

This is basically a log-linear model with all 2-way interactions.

Proof for the dichotomous, Joe & Liu (1996); for other cases, Anderson, Li & Vermunt (2007), and Anderson, Verkuilen & Peyton (in press).

Simplifying the Model

For each pair of items, there is a $(J \times L)$ matrix of ψ's,

$$\boldsymbol{\Psi}_{i|k} = \begin{pmatrix} \psi_{i1|k1} & \psi_{i1|k2} & \cdots & \psi_{i1|kL} \\ \psi_{i2|k1} & \psi_{i2|k2} & \cdots & \psi_{i2|kL} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{iJ|k1} & \psi_{iJ|k2} & \cdots & \psi_{iJ|kL} \end{pmatrix}$$

The model can be simplified by considering lower rank decompositions:

$$\boldsymbol{\Psi}_{i|k} = \boldsymbol{N}_i^{[ik]} \boldsymbol{\Sigma}^{[ik]} \boldsymbol{N}_k^{[ik]\prime} \qquad \text{where} \quad \boldsymbol{\Sigma}^{[ik]} \text{ is diagonal.}$$

However, here we'll mostly consider those of the form $\boldsymbol{\Psi}_{i|k} = \boldsymbol{N}_i \boldsymbol{\Sigma} \boldsymbol{N}_k'$ where $\boldsymbol{\Sigma}$ is not necessarily diagonal and

$$\underbrace{\boldsymbol{N}_i}_{(J \times M)} = \begin{pmatrix} \nu_{i11} & \nu_{i12} & \cdots & \nu_{i1M} \\ \nu_{i21} & \nu_{i22} & \cdots & \nu_{i2M} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_{iJ1} & \nu_{iJ2} & \cdots & \nu_{iJM} \end{pmatrix}$$

Special Case $\#1$: $M = 1$

$$\boldsymbol{\Psi}_{i|k} = \boldsymbol{\nu}_{i1}\sigma_{11}\boldsymbol{\nu}'_{k1} = \{\sigma_{11}\nu_{ij1}\nu_{k\ell 1}\}$$

The conditional logistic regression model for each item i is

$$\begin{aligned} P(Y_i = j | y_{k\ell}, k \neq i) &= \frac{\exp(\lambda_{ij} + \nu_{ij1}(\sigma_{11}\sum_{k\neq i}\nu_{k\ell 1}))}{\sum_h \exp(\lambda_{ih} + \nu_{ih1}(\sigma_{11}\sum_{k\neq i}\nu_{k\ell 1}))} \\ &= \frac{\exp(b_{ij} + a_{ij}\tilde{\theta})}{\sum_h \exp(b_{ih} + a_{ih}\tilde{\theta})} \end{aligned}$$

- $\lambda_{ij} = b_{ij}$ is an intercept or "difficulty" parameter.
- $\nu_{ij1} = a_{ij}$ is a slope or "discrimination" parameter.
- The predictor variable is a (weighted) rest-score: $\tilde{\theta} = \sigma_{11}\sum_{k\neq i}\nu_{k\ell 1}$. Justification, see Junker (1993), and Junker & Sijtsma (2000)
- Bock's nominal response model and all it's special cases.
- The LMA

$$P(\boldsymbol{y}) = \lambda + \sum_i \lambda_{ij} + \sigma_{11}\sum_i\sum_{k>i}\nu_{ij1}\nu_{k\ell 1}$$

Special Case $\#2$: M

$$\boldsymbol{\Psi}_{i|k} = \boldsymbol{N}_i \boldsymbol{\Sigma} \boldsymbol{N}_k^{'} = \left\{ \sum_m \sum_{m'} \nu_{ijm} \sigma_{mm'} \nu_{k\ell m'} \right\}$$

The conditional logistic regression model for each item i is

$$P(Y_i = j | y_{k\ell}, k \neq i) = \frac{\exp(\lambda_{ij} + \sum_m \nu_{ijm} (\sum_{m'} \sigma_{mm'} \sum_{k \neq i} \nu_{k\ell m'}))}{\sum_h \exp(\lambda_{ih} + \sum_m \nu_{ihm} (\sum_{m'} \sigma_{mm'} \sum_{k \neq i} \nu_{k\ell m'}))}$$

$$= \frac{\exp(b_{ij} + \sum_m a_{ijm} \tilde{\theta}_m)}{\sum_h \exp(b_{ih} + \sum_m a_{ihm} \tilde{\theta}_m)}$$

- $\nu_{ijm} = a_{ijm}$ is the slope or discrimination parameter for variable m.
- $\tilde{\theta}_m$ is predictor variable m or weighted sum of rest-scores or test-totals:

$$\tilde{\theta}_m = \sigma_{mm} \underbrace{\left(\sum_{k \neq i} \nu_{k\ell m}\right)}_{\text{rest-score } m} + \sum_{m' \neq m} \sigma_{mm'} \underbrace{\left(\sum_{k \neq i} \nu_{k\ell m'}\right)}_{\text{rest-score or test-total } m'} .$$

- Multidimensional compensatory IRT model for polytomous items.

Recent Developments: LMA as IRT Models

- Anderson, Li & Vermunt (2007):
  - Models in the Rasch family—dichotomous and polytomous items, uni– and multi–dimensional latent variables.
  - Pseudo-likelihood estimation.
- Anderson & Yu (2007):
  - Dichotomous items, 1 underlying latent variable.
  - Different underlying alterative marginal distributions of the latent variable.
- Anderson, Verkuilen & Peyton (in press):
  - Multicategory items and two latent variables (also 3 latent variable and higher order models, but these aren't in the paper).
  - Covariate that influenced the choice of the "don't know" response option (i.e., instructions). The covariate came in from the "left".
  - Hybrid model.
  - Equality restrictions on some parameters.
  - Models fit using SAS/NLP.

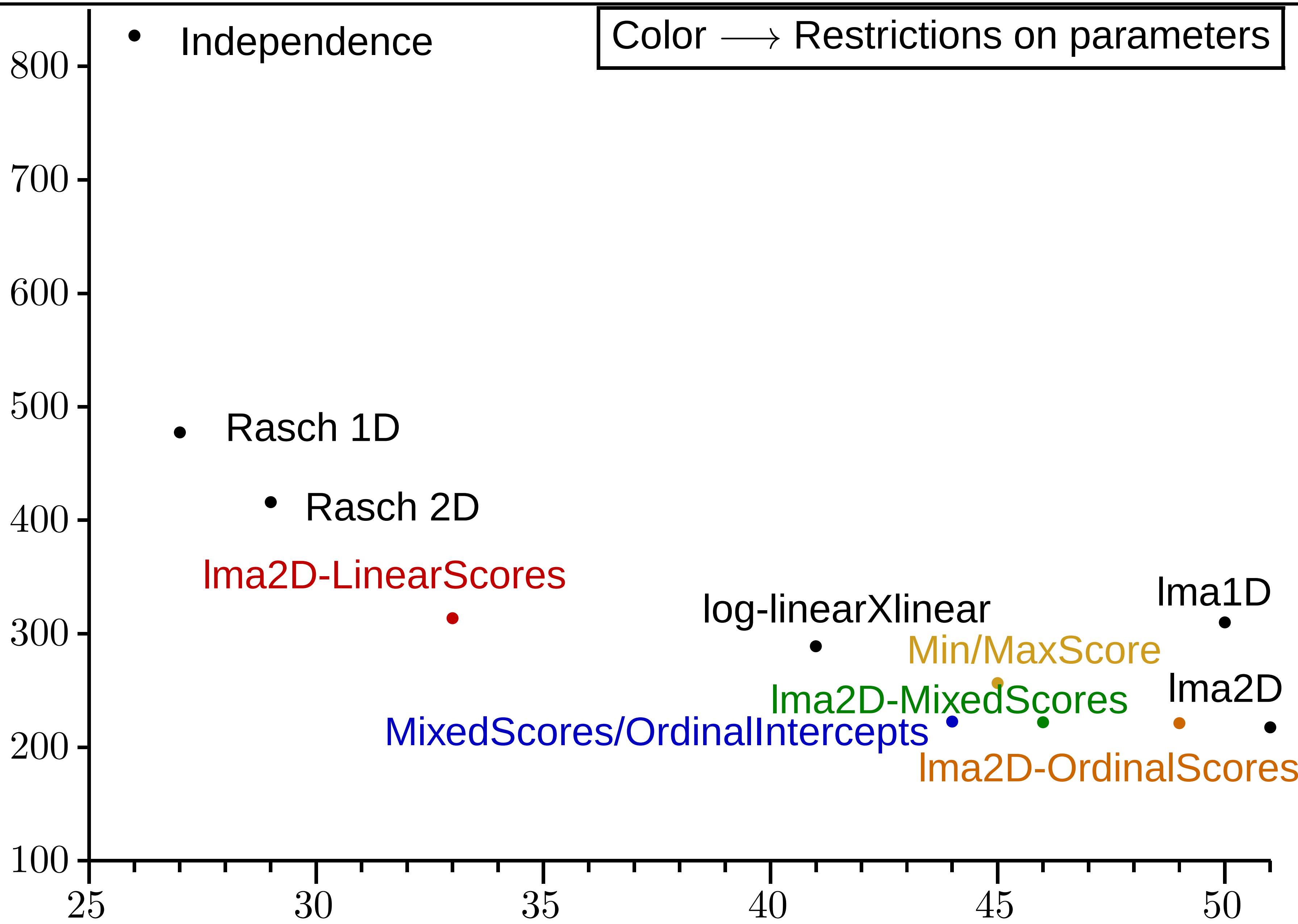
Common and Novel IRT Models as LMAs

	Restrictions on LMA Parameters	
Common models	λ_{ij} (b_{ij})	ν_{ijm} (a_{ijm})
2PL	none	none
Nominal response model	none	none
Multidimensional compensatory	none	none
Rasch family	none	input/fixed (ordered)
Graded $\times$ Nominal response	none	ordinal

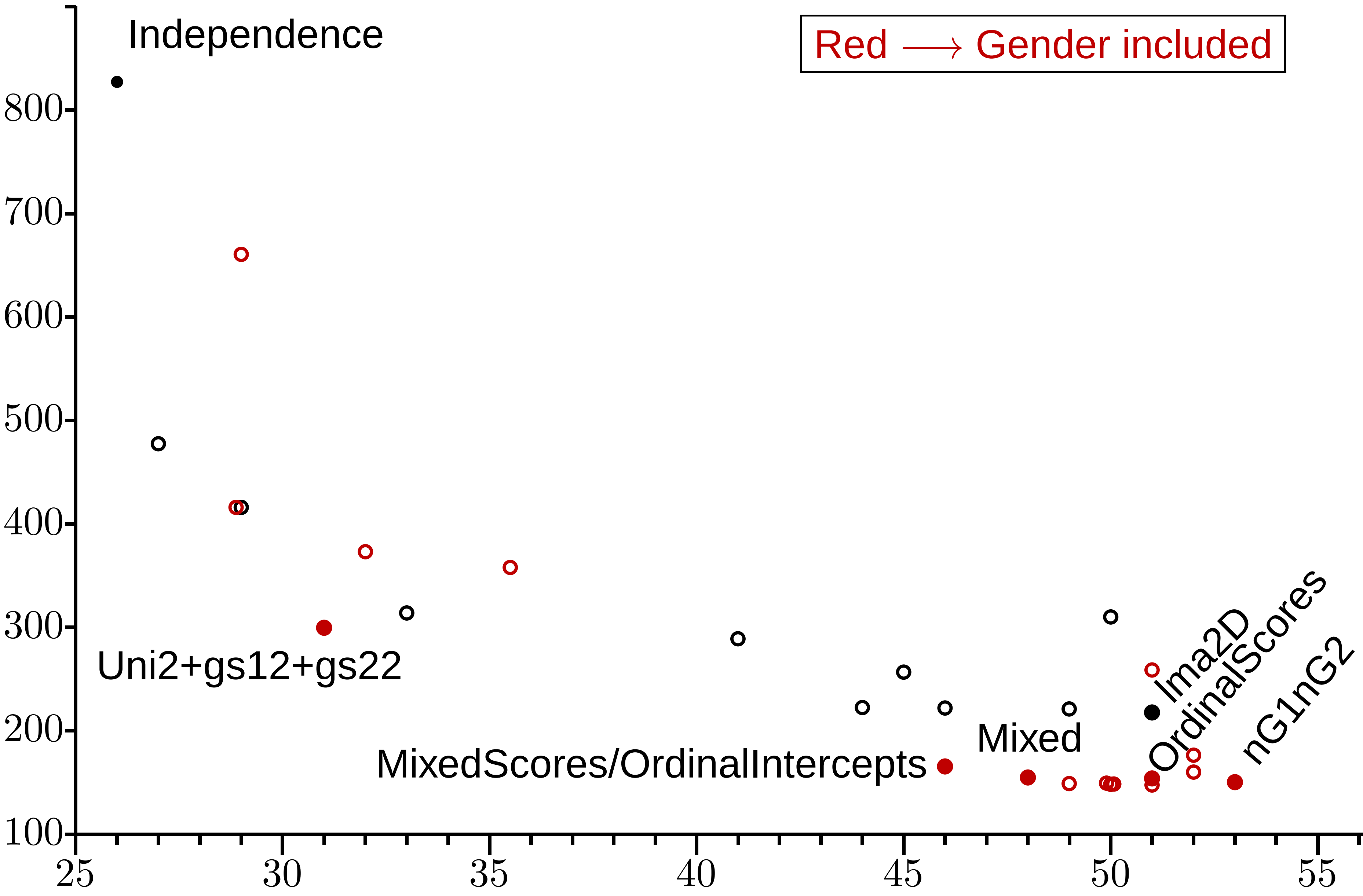
Novel ones can be created by modeling or placing restrictions on location parameters (i.e., λ_{ij}), category scores (i.e., ν_{ijm}), and/or σ_{mm}s:

- Input/fixed.
- Equality.
- Ordinal.
- Linear functions (e.g., $\nu_{ij} = \omega_i x_{ij}$ where $x_{ij} = 0, 1, \ldots J$ or any values).
- Set minimum and maximum (e.g., 0 and 1) and estimate scores in between.
- Model σ_{mm} (e.g., $\sigma_{mm} = \sigma^*_{mm} + \beta_m x$ or $\sigma_{mm} = \sigma^*_{mm}\beta_m x$).
- Other: e.g., $\nu_{ij} = \omega_i \nu^*_{ij}$ where $\sum_j \nu^{*2}_{ij} = 1$ or set min and max ν^*_{ij}.

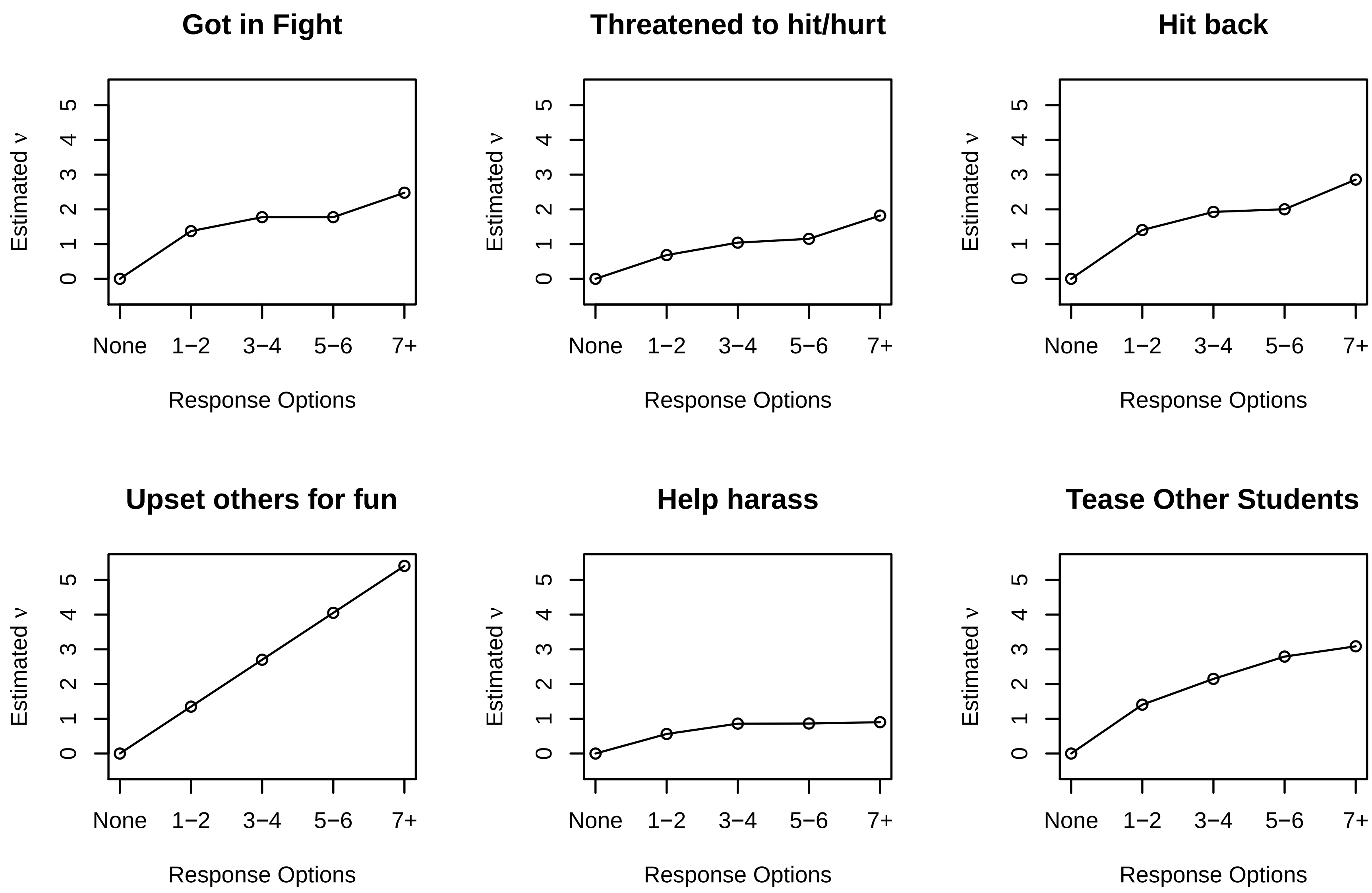
-2lnlike versus Number Parameters



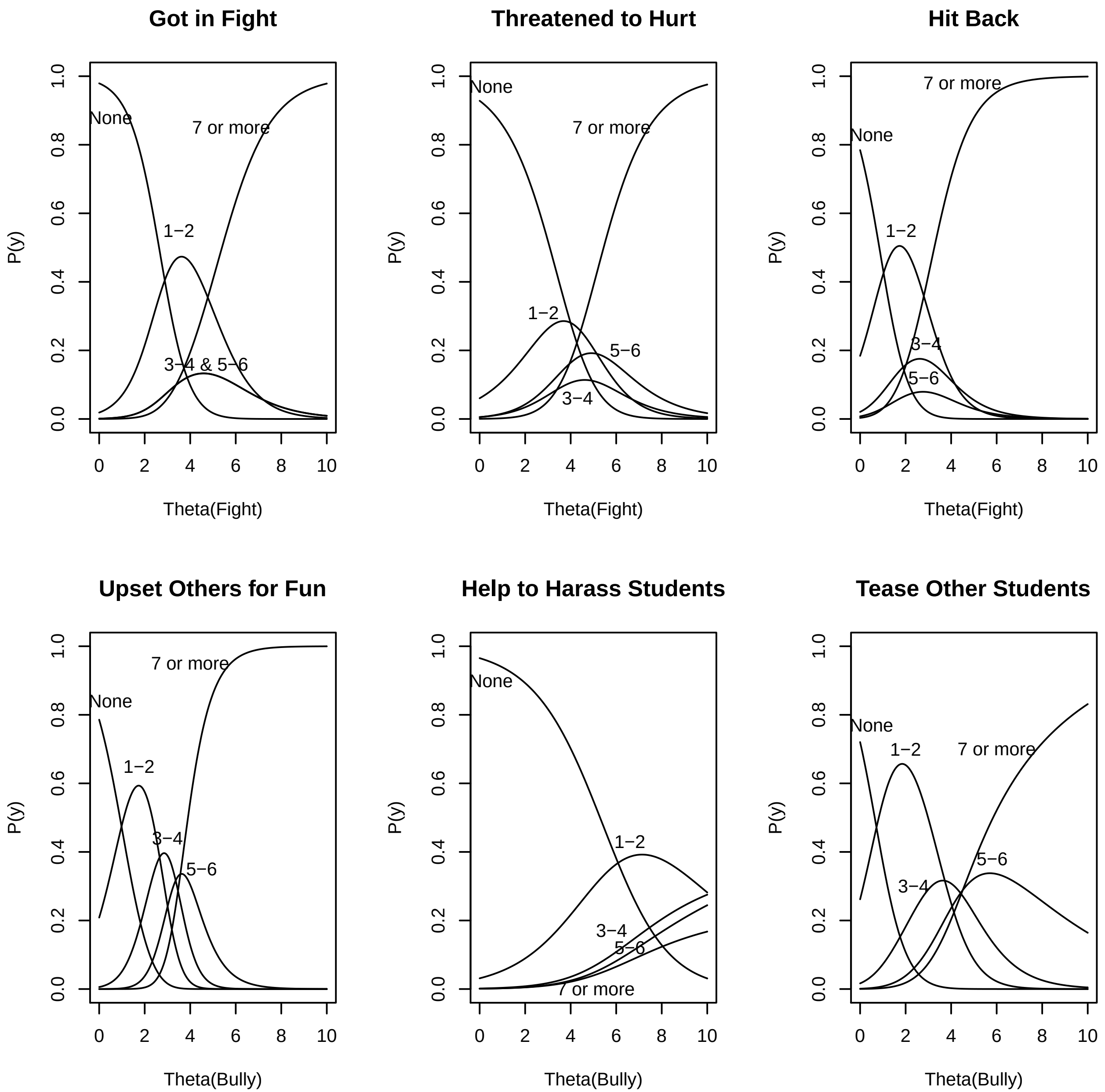
-2lnlike versus Number Parameters w/ Gender



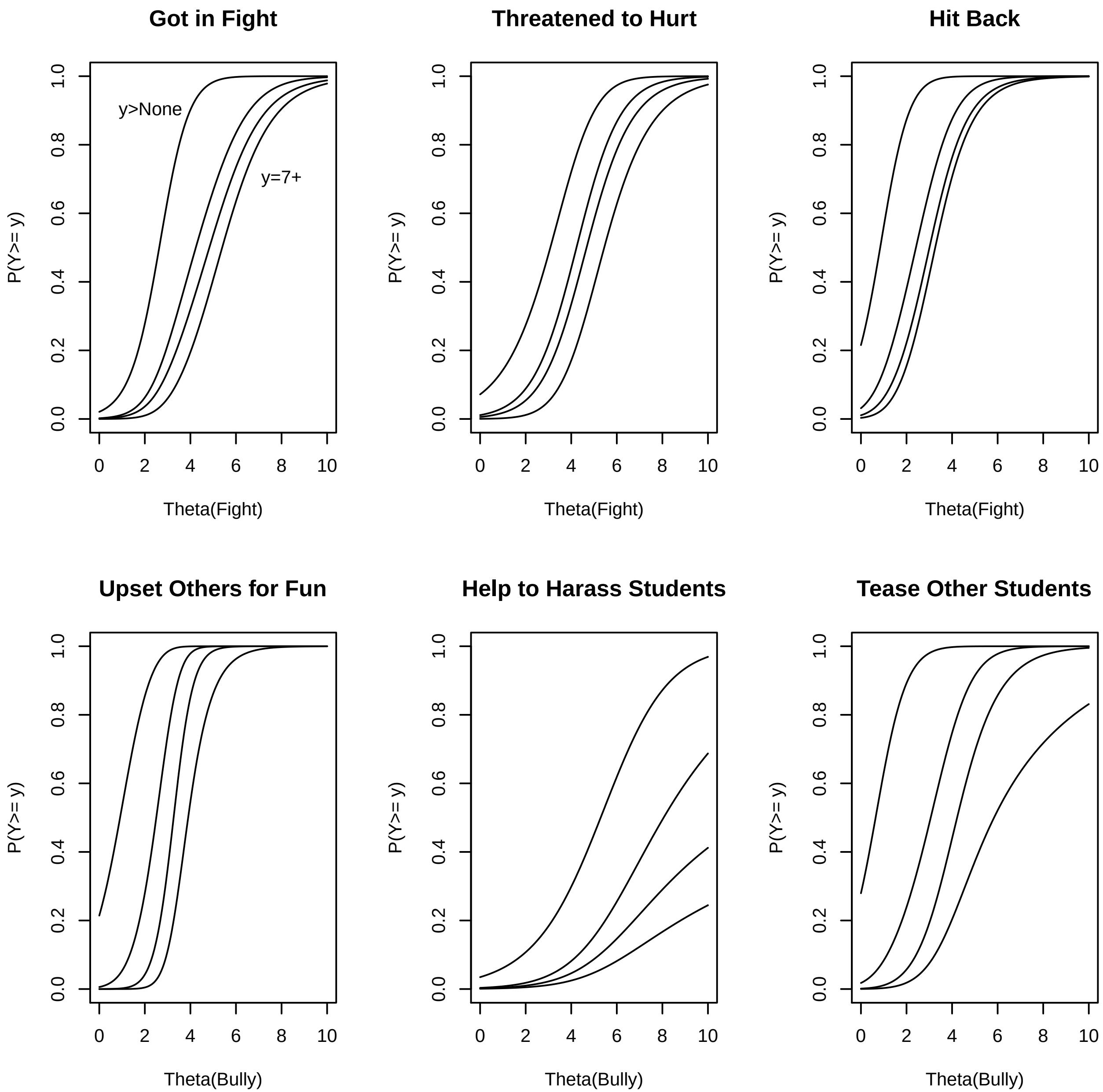
Estimated Item Scale Values



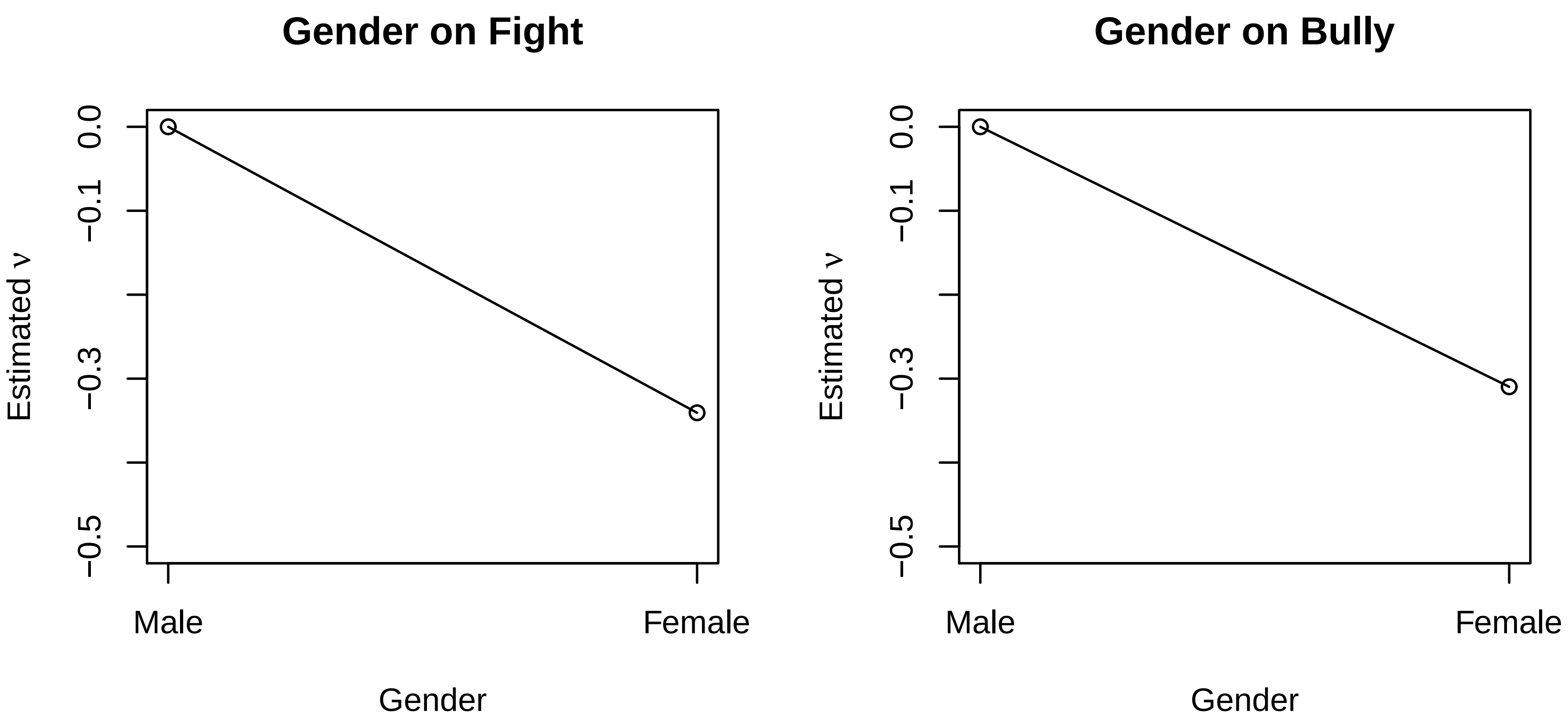
Item Characteristic Curves



Item Cumulative Probability Curves



Estimated Scale Values for Gender



Example Illustrated. . .

Noteworthy in today's example for multicategory items:

- Marginal distribution of traits are very skewed.
- Ordinal restrictions on responses options (and linear transformations).
- Ordinal restrictions on intercepts ("difficulties" or location parameters).
- Covariate came in from the "right" (i.e., helps model underlying latent variable).
- The scale values for categories of the 3 three bully items are nearly identical from a uni-dimensional LMA model fit to all 9 bully items (and different estimation methods). Correlations $> .99$.

  The Importance of this: Illustrates that major criticisms of conditional models do not hold up for LMA models as latent variable models:
  - Models parameters are essentially the same.
  - No interpretational difficulty.

Estimation Developments

The major problem is the size of a table (i.e., number of items/categories), but not the number of latent variables.

- The largest problem that I've successfully fit using ℓEM (Vermunt, 1997) is $2^{12} = 4096$ response patterns.
- Bayesian methods for the $RC(M)$ association model for 2-way tables (Iliopoulos, Kateri, & Ntzoufras, 2007; Iliopoulos & Kateri, 2009)
- Models can be fit to data using SAS/NLP (probably also R and MatLab using their optimization capabilities.). Although this approach can fit larger numbers of items/categories than ℓEM, it is still somewhat limited.
- Models in the Rasch family can be fit by pseudo-likelihood estimation in any program that can fit conditional logistic regression models (Anderson, Li & Vermunt, 2007) and can include covariates.

  No limit hit (yet) in terms of number of items/categories or number of latent variables.

- For models with estimated category scores, an experimental algorithm.

Experimental Algorithm

For models with estimated category scores, an experimental algorithm that iteratively fits conditional logistic regressions using MLE to estimate scale values and a pseudo-likelihood step to estimate the association parameters.

- Takes advantage of conditional specification of models.
- Applications to data sets yield nearly identical estimates as MLE.
- In simulation studies, the algorithm yields parameters estimates nearly identical (up to linear transformation) of the parameters used to simulate data (by some IRT model).
- Converges relatively quickly.
- I haven’t hit a limit in terms of number of items/categories.

Areas for Future Research

- Estimation.
- Applications, especially those with multiple latent variables (e.g., testlets, Q-matrix).
- Further comparisons with traditional IRT methods (i.e., estimation of item parameters and individuals' values on latent variables).
- Handling missing data
- Addition of random effects
- Multidimensional, partially-compensatory models $\longrightarrow$ leads to higher-way interactions in model for the data where the higher-way interactions have higher-way decompositions (Tucker 3-mode and other higher-way type decompositions).

Downloads

Will be able to download SAS/NLP programs used in this talk and various papers from

http://faculty.ed.uiuc.edu/cja/homepage/software_index.html

and slides from

http://faculty.ed.uiuc.edu/cja/homepage

A: Who has the final responsibility to decide if a law is constitutional or not?
President (9.0%), Congress (27.6%), Supreme Court (57.9%), Don't know (5.5%)

B: Whose responsibility is it to nominate judges to the Federal courts?
President (50.8%), Congress (15.6%), Supreme Court (25.5%), Don't know (8.2%)

C: Do you happen to know which party has the most members in the House of Representatives in Washington? Republicans (70.3%), Democrats (16.0%), Don't know (13.7%)

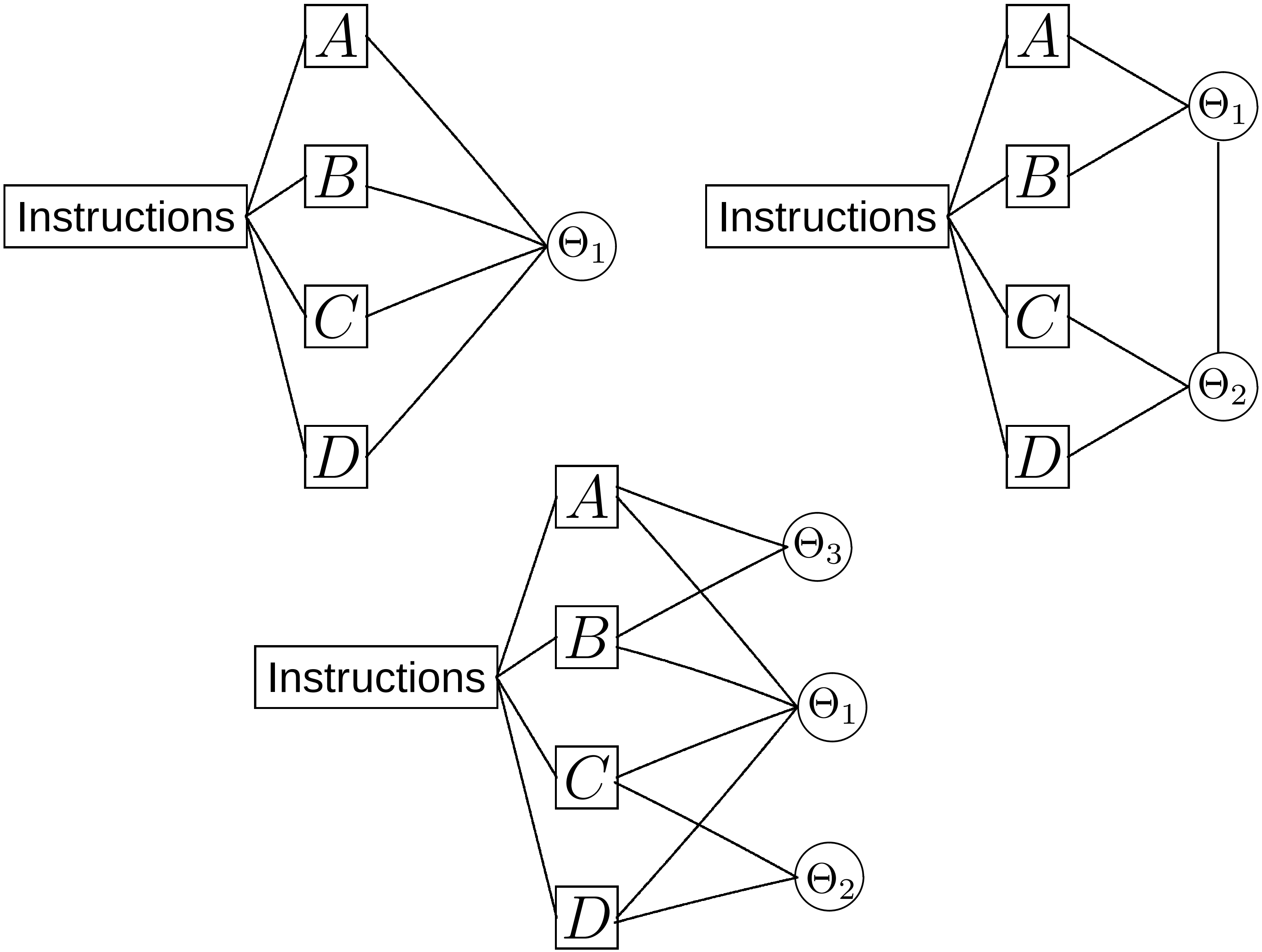
D: Do you happen to know which party has the most members in the U.S. Senate? Republicans (62.8%), Democrats (19.0%), Don't know (18.2%)

Anderson, C.J., Verkuilen, J.V., & Peyton, B.L. (in press). *Journal of Educational and Behavioral Statistics*.

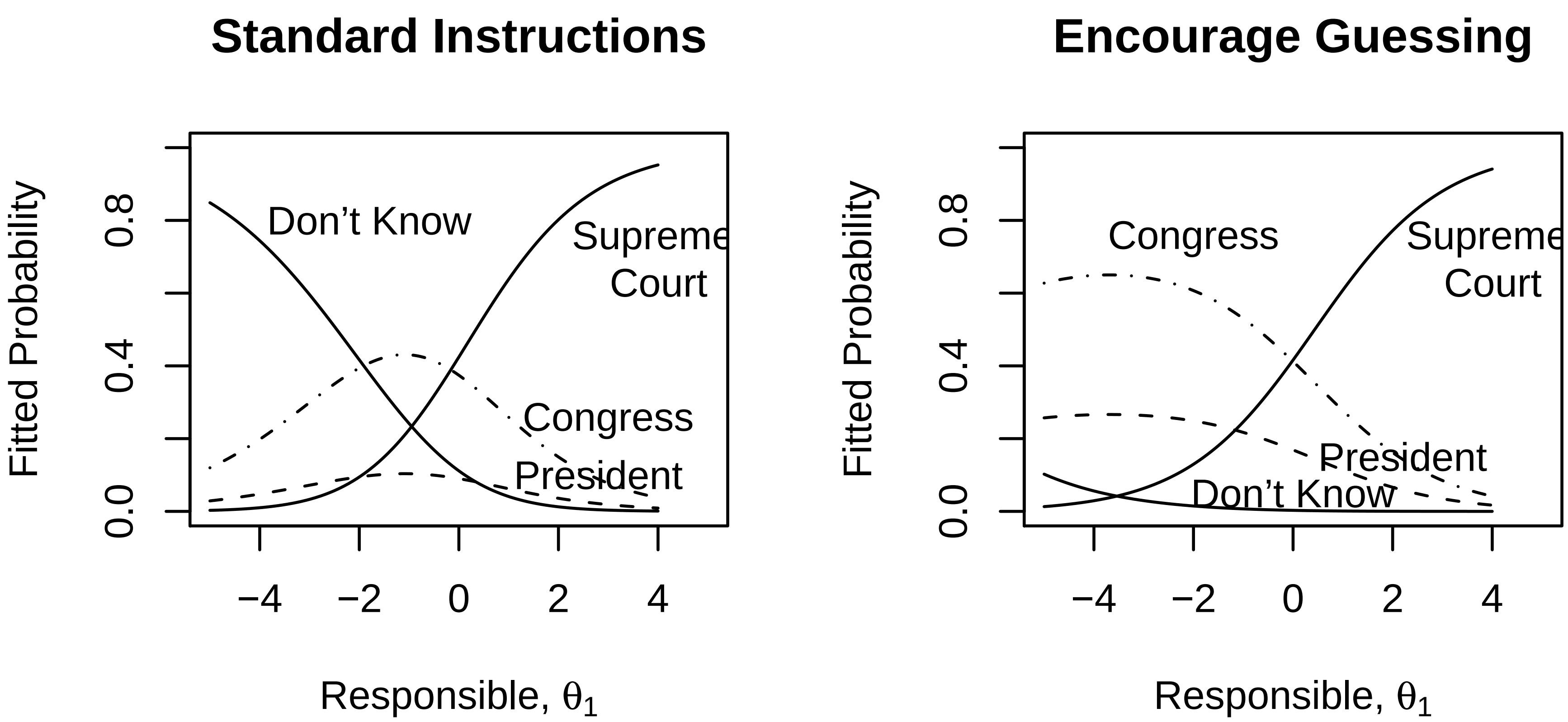
Challanges for Analysis of ANES Data

- Ordering of response the options unknown.
- Dealing with "Don't Know".
- Different number of response options.
- Scoring of responses needed.
- Latent variable structure unknown.
  - One dominant underlying dimension of "Knowledge."
  - Two correlated latent variables: Structure of political system (items A & B) and Party in Power (items C & D)
- How do the instructions given to respondents affect all of this?

ANES Graphs


Instructions
A
B
C
D
Θ_1
Instructions
A
B
C
D
Θ_1
Θ_2
Instructions
A
B
C
D
Θ_3
Θ_1
Θ_2

Effect of Instructions: Constitutionality of Laws



Andersen, E.B. (1995). The derivation of polytomous Rasch models. In G.H. Fischer & I.W. Molenaar (Eds) *Rasch Models: Foundations, Recent Developments, and Applications*, (pp. 271–291). NY: Springer.

Anderson, C.J., & Böckenholt, U. (2000). RC regression models for polytomous variables. *Psychometrika*, *65*, 479–509.

Anderson, C.J., Li, Z., Vermunt, J.K. (2007). Estimation of models in a Rasch family for polytomous items and multiple latent variables. *Journal of Statistical Software*, *20*. DOI: http://www.jstatsoft.org/v20/i06/v20i06.pdf.

Anderson, C.J., Verkuilen, J., Peyton, B. (in press). Modeling polytomous item responses using simultaneously estimated multinomial logistic regression models. *Journal of Educational and Behavioral Statistics*.

Anderson, C.J., & Vermunt, J.K. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology*, *30*, 81–121.

Anderson, C.J. & Yu, H.T. (2007). Log-multiplicative association models as item response models. *Psychometrika,* *72*, 5–23.

Becker, M. (1989). On the bivariate normal distribution in multivariate contingency tables. *Statistics and Probability Letters, 8*, 435–440.

de Rooij, M. (2008). The analysis of change, Newton's law of gravity and association models. *Journal of the Royal Statistical Society, Series A*, *171*, 137–157.

de Rooij, M., & Heiser, W. (2005). Graphical representations and odds ratios in a distance-association model for the analysis of cross-classified data. *Psychometrika, 70*, 1–24.

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, **74**, 537–552.

Goodman, L.A. (1985). The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *The Annals of Statistics, 13*, 10–69.

Goodman, L.A. (1986). Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review, 54*, 234–309.

Joe, H., & Liu, Y. (1996). A model for multivariate binary response with covariates based on conditionally specified logistic regressions. *Statistics & Probability Letters, 31*, 113–120.

Junker, B.W. (1993). Conditional association, essential independence and monotone unidimensional item response models. *Annals of Statistics, 21*, 1359-1378.

Junker, B.W., & Sijtsma, K. (2000). Latent and manifest monotonicity in item response models. *Applied Psychological Measurement, 24*, 65–81.

Li, Z. (2010). *Log-linear Models as Item Response Models*. University of Illinois, Urbana-Champaign. Unpublished Ph.D dissertation.

Takane, Y. (1987). Analysis of contingency tables by ideal point discriminant analysis. *Psychometrika*, *52*, 493–513.

Tettegah, S., & Anderson, C. J. (2007). Pre-service teachers' empathy and cognitions: Statistical analysis of text data by graphical models. *Contemporary Educational Psychology*, *32*, 48–82.

Vermunt, J.K. (1997). ℓ_{EM}*: A General Program for the analysis of categorical data.* [Computer software manual]. Tilburg, The Netherlands. Retrieved from http://www.uvt.nl /faculteiten/fsw/organisatie/departementen/mto/software2.html.