Simultaneous Estimation of Multinomial Logistic Regression Models: Factor Analysis of Polytomous Item Response Data

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Introduction

ANES Data
Existing Approaches
Conditional Specification Approach
Extensions
Analysis of the ANES Data
Conclusion & Beyond

- Flexible latent variable models for multivariate, polytomous nominal and/or ordinal variables with collateral information.

- Observed discrete (multicategory) variables, e.g.,
  - Multiple choice test items.
  - Surveys items.
  - “Likert”.
  - Coded qualitative data.

- Collateral Information includes observed covariates (discrete or continuous) that are measures of individuals, items, response options, latent variables, as well as specific hypotheses about items and response options.

- (Multiple) Unobserved continuous variables.

American National Elections Pilot (1998)

ANES Data
- American National Elections Pilot (1998)
- Challanges for Analysis of ANES Data

Existing Approaches

Conditional Specification Approach

Extensions

Analysis of the ANES Data

Conclusion & Beyond

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. (accepted). http://faculty.ed.uiuc.edu/cja/homepage → "Publications".

Challanges for Analysis of ANES Data

ANES Data
- American National Elections Pilot (1998)
- Challanges for Analysis of ANES Data

Existing Approaches

Conditional Specification Approach

Extensions

Analysis of the ANES Data

Conclusion & Beyond

- 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?

Major Existing Approaches

ANES Data

Existing Approaches
- Major Existing Approaches

Conditional Specification Approach

Extensions

Analysis of the ANES Data

Conclusion & Beyond

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 (e.g., Bock's multinomial response model)
  - 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)

Conditional Specification Approach

ANES Data
Existing Approaches
Conditional Specification Approach
● Conditional Specification Approach
● Conditional Specification Approach: ANES
● The Joint Distribution
● Log-multiplicative Association Models
Extensions
Analysis of the ANES Data
Conclusion & Beyond

or the Item Response Theory Approach

1. Bock's nominal response model, e.g.,

$$P(Y_i = j_i|\theta) = \frac{\exp[\lambda_{ij_i} + \nu_{ij_i}\theta]}{\sum_{h_i} \exp[\lambda_{ih_i} + \nu_{ih_i}\theta]}.$$

2. If we knew or had estimates of the ν_{ij_i}'s for all the items, then $\sum_i \nu_{ij_i}$ would be sufficient for θ (Andersen, 1995).

3. Suppose we know all the ν's except for item i's, then we could replace θ with a "rest-score",

$$\tilde{\theta}_{-i} = \phi \sum_{k \neq i} \nu_{kj_k}.$$

- This gives us a multinomial logistic regression model where the predictor variable is a rest-score.
- Justification and precedence comes from item response theory and classical test theory (Junker, 1993; Junker & Sijtsma, 2000).

Conditional Specification Approach: ANES

ANES Data

Existing Approaches

Conditional Specification Approach
- Conditional Specification Approach
- Conditional Specification Approach: ANES
- The Joint Distribution
- Log-multiplicative Association Models

Extensions

Analysis of the ANES Data

Conclusion & Beyond

$$\begin{aligned}
\text{Laws:}\quad & P(A = j_1|\tilde{\theta}_A) &&= \exp[\lambda^A_{j_1} + \nu^A_{j_1}\phi(\nu^B_{j_2} + \nu^C_{j_3} + \nu^D_{j_4})]\kappa_A \\
\text{Judges:}\quad & P(B = j_2|\tilde{\theta}_B) &&= \exp[\lambda^B_{j_2} + \nu^B_{j_2}\phi(\nu^A_{j_1} + \nu^C_{j_3} + \nu^D_{j_4})]\kappa_B \\
\text{House:}\quad & P(C = j_3|\tilde{\theta}_C) &&= \exp[\lambda^C_{j_3} + \nu^C_{j_3}\phi(\nu^A_{j_1} + \nu^B_{j_2} + \nu^D_{j_4})]\kappa_C \\
\text{Senate:}\quad & P(D = j_4|\tilde{\theta}_D) &&= \exp[\lambda^D_{j_4} + \nu^D_{j_4}\phi^2(\nu^A_{j_1} + \nu^B_{j_2} + \nu^C_{j_3})]\kappa_D
\end{aligned}$$

The κ's ensure that probabilities over response options sum to 1 for each item.

Compatibility Conditions are needed to ensure this set is consistent with some joint distribution:

> When variable i is the response variable, the term relating variable i' to i must be the same as the term relating i to i' when i' is the response.

The proof is in paper for all models discussed in this talk.

The Joint Distribution

ANES Data

Existing Approaches

Conditional Specification Approach

- Conditional Specification Approach
- Conditional Specification Approach: ANES
- The Joint Distribution
- Log-multiplicative Association Models

Extensions

Analysis of the ANES Data

Conclusion & Beyond

A set of multinomial logistic regression models that meet the compatibility conditions uniquely implies a log-multiplicative association model for the joint distribution.

For the ANES data:

$$\begin{aligned}\log(P(\boldsymbol{y})) \quad &= \quad \lambda + \lambda^A_{j_1} + \lambda^B_{j_2} + \lambda^C_{j_3} + \lambda^D_{j_4} \\ &\quad +\phi\left(\nu^A_{j_1}\nu^B_{j_2} + \nu^A_{j_1}\nu^C_{j_3} + \nu^A_{j_1}\nu^D_{j_4} + \nu^B_{j_2}\nu^C_{j_3}\right. \\ &\quad \left. +\nu^B_{j_2}\nu^D_{j_4} + \nu^C_{j_3}\nu^D_{j_4}\right).\end{aligned}$$

A log-multiplicative association model.

Log-multiplicative Association Models

ANES Data

Existing Approaches

Conditional Specification Approach

- Conditional Specification Approach
- Conditional Specification Approach: ANES
- The Joint Distribution
- Log-multiplicative Association Models

Extensions

Analysis of the ANES Data

Conclusion & Beyond

- Are special cases of Poisson regression models (i.e., log-linear models).
- Are generalizations of Goodman's (1979, 1986) multidimensional row-column or "$RC(M)$" association models for 2-way tables.
- Are implied by underlying multivariate normal distribution (Goodman, 1979; Becker, 1989; others).
- Can be derived from a distance based model (de Rooij & Heiser, 2005), a generalization of Newton's Law of Gravity.
- 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, accepted).

Extensions

ANES Data

Existing Approaches

Conditional Specification Approach

Extensions
- Extensions
- Multiple (correlated) latent variables
- Joint Distribution For M latent variables
- Restrictions on Parameters
- Hybrid Item Response Models
- Covariates

Analysis of the ANES Data

Conclusion & Beyond

To be discussed here...

1. M Multiple (correlated) latent variables.
2. Restrictions on parameters
   - Equality
   - Scaling
   - Other
3. Adding covariates
   - Information about response options.
   - Information about instructions (treatment condition).

Multiple (correlated) latent variables

ANES Data
Existing Approaches
Conditional Specification Approach
Extensions
● Extensions
● Multiple (correlated) latent variables
● Joint Distribution For M latent variables
● Restrictions on Parameters
● Hybrid Item Response Models
● Covariates
Analysis of the ANES Data
Conclusion & Beyond

$$\tilde{\theta}_{m,-i} = \phi_{mm} \sum_{k \neq i} \nu_{kj_k m} + \sum_{m' \neq m} \phi_{m'm} \left(\sum_k \nu_{kj_k m'} \right)$$

and the conditional multinomial model is

$$P(Y_i = j_i | \tilde{\theta}_{m,-i}, \forall m) = \frac{\exp[\lambda_{ij_i} + \sum_m \nu_{ij_i m} \tilde{\theta}_{m,-i}]}{\sum_{h_i} \exp[\lambda_{ih_i} + \sum_m \nu_{ih_i m} \tilde{\theta}_{m,-i}]}.$$

- If there is no direct relationship between an item and a latent variable, then the corresponding $\nu_{ij_i m} = 0$.
- The compatibility conditions are

$$\phi_{mm'} \nu_{ij_i m} \nu_{kj_k m'} = \phi_{m'm} \nu_{kj_k m'} \nu_{ij_i m}$$

for all items, categories and latent variables.

i.e., $\phi_{mm'} = \phi_{m'm}$.

Joint Distribution For M latent variables

ANES Data

Existing Approaches

Conditional Specification Approach

Extensions
- Extensions
- Multiple (correlated) latent variables
- Joint Distribution For M latent variables
- Restrictions on Parameters
- Hybrid Item Response Models
- Covariates

Analysis of the ANES Data

Conclusion & Beyond

The terms in the model include

1. A λ to ensure that probabilities summed over all possible response patterns equal 1.
2. All marginal effects terms λ_{ij_i} (i.e., the intercepts in the multinomials).
3. All multiplicative terms $\phi_{mm'}\nu_{ij_im}\nu_{kj_km}$ representing the relationship between pairs of items in the set of models defined by the multinomials.

Applying these rules yields

$$P(\boldsymbol{y}) = \exp\left[\lambda + \sum_i \lambda_{ij_i} + \sum_i \sum_{k>i} \sum_m \sum_{m'} \phi_{mm'}\nu_{ij_im}\nu_{kj_km'}.\right]$$

Restrictions on Parameters

ANES Data

Existing Approaches

Conditional Specification Approach

Extensions

- Extensions
- Multiple (correlated) latent variables
- Joint Distribution For M latent variables
- Restrictions on Parameters
- Hybrid Item Response Models
- Covariates

Analysis of the ANES Data

Conclusion & Beyond

- Equality and ordinal restrictions on category scores are straight forward.

- For identification, need to set the
  - Location the λ_{ij_i} and category scores ν_{ij_i}, e.g.,

$$\sum_{j_i} \lambda_{ij_i} = \sum_{j_i} \nu_{ij_i} = 0$$

  - Scale of one set of ν_{ij_i} per latent variable, e.g.,

$$\sum_{j_i} \nu_{ij_i}^2 = 1$$

  - The category scale can be re-expressed as

$$\nu_{ij_i} = \omega_i \nu^*_{ij_i}$$

where $\omega_i = \sqrt{\sum_{j_i} \nu_{ij_i}^2}$.

Hybrid Item Response Models

ANES Data
Existing Approaches
Conditional Specification Approach
Extensions
- Extensions
- Multiple (correlated) latent variables
- Joint Distribution For M latent variables
- Restrictions on Parameters
- Hybrid Item Response Models
- Covariates

Analysis of the ANES Data
Conclusion & Beyond

- The category scale can be re-expressed as

$$\nu_{ij_i} = \omega_i \nu^*_{ij_i}$$

  where $\omega_i = \sqrt{\sum_{j_i} \nu^2_{ij_i}}$.
- Set some (or all) $\omega_i = 1$ yields a "hybrid" item response model where
  - Each item is equally strongly related to the latent variable (Rasch-like).
  - The category scores $\nu^*_{ij_i}$ carry category specific information (Bock NRM-like).
- The restriction of equal $\omega_i (= 1)$ for items can be imposed by placing scaling restriction on the category scores; that is,

$$\sum_{j_i} \nu^2_{ij_i} = 1$$

  on categories scores for more items than what is needed for identification.

Covariates

ANES Data

Existing Approaches

Conditional Specification Approach

Extensions
- Extensions
- Multiple (correlated) latent variables
- Joint Distribution For M latent variables
- Restrictions on Parameters
- Hybrid Item Response Models
- Covariates

Analysis of the ANES Data

Conclusion & Beyond

Covariates can be added to the model by creating or modifying sub-models for the intercepts and/or latent variable(s).

- Possible models for the intercept:

$$\lambda_{ij_i} = \lambda^*_{ij_i} + \sum_q \beta_{ij_i q} x_q$$

$$\lambda_{ij_i} = \lambda^*_{ij_i} + \sum_q \beta_{ij_i q} x_{iq}$$

$$\lambda_{ij_i} = \lambda^*_{ij_i} + \sum_q \beta_q x_{ij_i q}$$

- A possible model for a latent variable (i.e., covariate adds information about a person's value on the latent variable):

$$\tilde{\theta}_{-i} = \phi \left(\sum_{k \neq i} \nu_{kj_k} + \sum_q \eta_q x_q \right) .$$

Analysis of the ANES Data

ANES Data

Existing Approaches

Conditional Specification Approach

Extensions

Analysis of the ANES Data
- Analysis of the ANES Data
- Possible Graphs for the ANES Data
- Analysis of the ANES Data
- Scoring and "Don't Know"
- Effect of Instructions: Constitutionality of Laws
- Single Measure of Political Knowledge

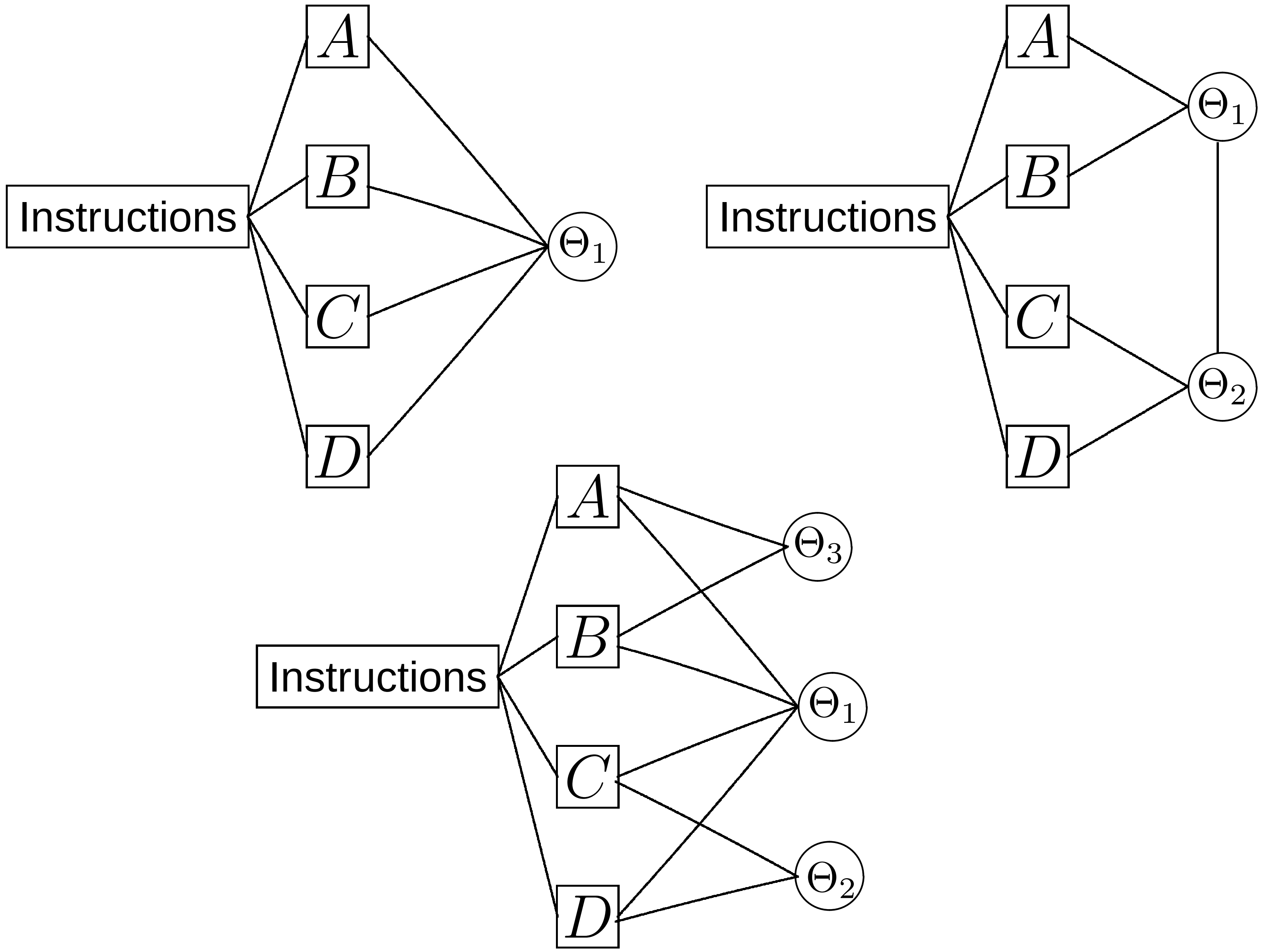
Conclusion & Beyond

1. Hypotheses regarding "Don't know":
   - Mondak (2001): "Don't know" responders might have intermediate states of knowledge but be unwilling to venture a substantive response and thus be scored wrong even though they knew more than responders who answer incorrectly.
   - "Don't know" responders have less knowledge than those who respond with a substantive but incorrect answer.
   - "Don't know" responses are simply not comparable to substantive responses, be they correct or incorrect.
2. Instructions given to respondents would effect the way they responded. Treatment condition:
   - Standard or control group ($n = 606$)
   - Instructions that encouraged respondents to guess if they were not sure of the correct answer ($n = 570$).
3. Goal: Measure political knowledge.

Possible Graphs for the ANES Data

ANES Data
Existing Approaches
Conditional Specification Approach
Extensions
Analysis of the ANES Data
- Analysis of the ANES Data
- Possible Graphs for the ANES Data
- Analysis of the ANES Data
- Scoring and “Don’t Know”
- Effect of Instructions: Constitutionality of Laws
- Single Measure of Political Knowledge

Conclusion & Beyond



Analysis of the ANES Data

ANES Data

Existing Approaches

Conditional Specification Approach

Extensions

Analysis of the ANES Data
- Analysis of the ANES Data
- Possible Graphs for the ANES Data
- Analysis of the ANES Data
- Scoring and "Don't Know"
- Effect of Instructions: Constitutionality of Laws
- Single Measure of Political Knowledge

Conclusion & Beyond

Fit statistics for models fit to the cross-classification of instructions and responses to the four ANES items.

Model	# par	G^2	BIC	Percent account
Log-linear Models				
Independence	11	1533.02	-418.27	—
All 2-way	58	189.66	-1429.34	88%
Log-multiplicative association models (LMA)				
One dimension	31	418.22	-1391.67	73%
Two dimensions	32	312.88	-1489.94	80%
2D Hybrid	30	313.12	-1503.84	80%
2D Hybrid with $\nu_{11}^{A} = \nu_{21}^{A}$	29	316.79	-1507.23	79%
Three dimensional LMA	38	227.54	-1532.86	85%
3D Hybrid	36	229.69	-1544.84	85%
3D Hybrid w/ $\nu_{11}^{A} = \nu_{21}^{A}$	35	230.33	-1551.28	85%
3D Hybrid w/ $\nu_{11}^{A} = \nu_{21}^{A}$, $\nu_{j_3 2}^{C} = \nu_{j_3 2}^{D}$	34	230.34	-1558.33	85%

Scoring and "Don't Know"

ANES Data
Existing Approaches
Conditional Specification Approach
Extensions
Analysis of the ANES Data
- Analysis of the ANES Data
- Possible Graphs for the ANES Data
- Analysis of the ANES Data
- Scoring and "Don't Know"
- Effect of Instructions: Constitutionality of Laws
- Single Measure of Political Knowledge

Conclusion & Beyond

The estimated scale values/scores:

Item	Category	Parameter	Estimate
A: Determines	President	ν_{11}^{A}	-0.036
constitutionality	Congress	ν_{21}^{A}	-0.036
of a law	Supreme Ct.	ν_{31}^{A}	0.741
	Don't know	ν_{41}^{A}	-0.700
B: Responsible for	President	ν_{11}^{B}	0.775
nominating federal	Congress	ν_{21}^{B}	0.095
judges	Supreme Ct.	ν_{31}^{B}	-0.359
	Don't know	ν_{41}^{B}	-0.511
C: Holds majority	Republican	ν_{12}^{C}	0.549
in the House	Democrat	ν_{322}^{C}	0.249
	Don't know	ν_{32}^{C}	-0.798
D: Holds majority	Republican	ν_{12}^{D}	0.625
in the Senate	Democrat	ν_{22}^{D}	0.143
	Don't know	ν_{32}^{D}	-0.768

Effect of Instructions: Constitutionality of Laws

ANES Data

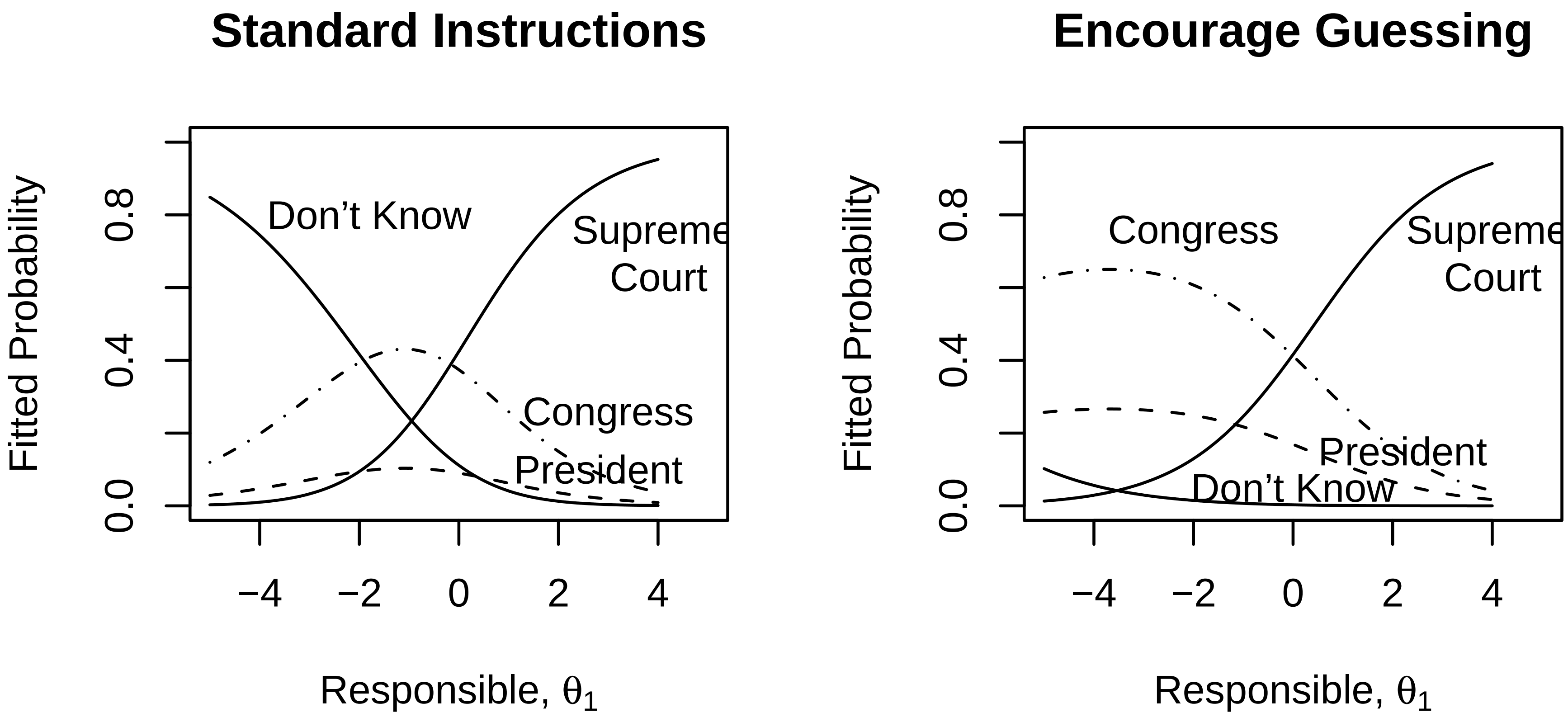
Existing Approaches

Conditional Specification Approach

Extensions

Analysis of the ANES Data
- Analysis of the ANES Data
- Possible Graphs for the ANES Data
- Analysis of the ANES Data
- Scoring and "Don't Know"
- Effect of Instructions: Constitutionality of Laws
- Single Measure of Political Knowledge

Conclusion & Beyond



Single Measure of Political Knowledge

ANES Data

Existing Approaches

Conditional Specification Approach

Extensions

Analysis of the ANES Data
- Analysis of the ANES Data
- Possible Graphs for the ANES Data
- Analysis of the ANES Data
- Scoring and “Don’t Know”
- Effect of Instructions: Constitutionality of Laws
- Single Measure of Political Knowledge

Conclusion & Beyond

and another look at the effect of Instructions

	Cronbach’s alpha (95% CI’s)			
Scoring Method	Control Group		Guessing Group	
Binary	.68	$(.64, .72)$	.58	$(.53, .63)$
LMA category scores	.75	$(.68, .82)$	.68	$(.60, .75)$

Comments:

- Scores from all the LMA models are all very similar.
- Model based scores lead to larger reliabilities.
- Less reliable measures when guessing is encouraged.

Conclusion & Beyond

ANES Data

Existing Approaches

Conditional Specification Approach

Extensions

Analysis of the ANES Data

Conclusion & Beyond

- Conclusion & Beyond
- Higher-Level Factors

- LMA models are effective tools for analysis and measurement of polytomous response data where latent variables may underling responses.

- Estimation that's more flexible and capable of larger problems.
  - ℓ_{EM} can handle about 2^{12}.
  - SAS PROC NLP can handle about 2^{20}.
  - Pseudolikelihood of models in the Rasch family can handle at least 5^{100}.
  - Experimental algorithm can handle at least 2^{100}.

- Some Additional Possibilities:
  - Other transformations of scale values, including ordinal and partial ordinal restrictions, spline, linear.
  - Higher level factors (i.e., lower rank decomposition of matrix of ϕ's).

Higher-Level Factors

ANES Data

Existing Approaches

Conditional Specification Approach

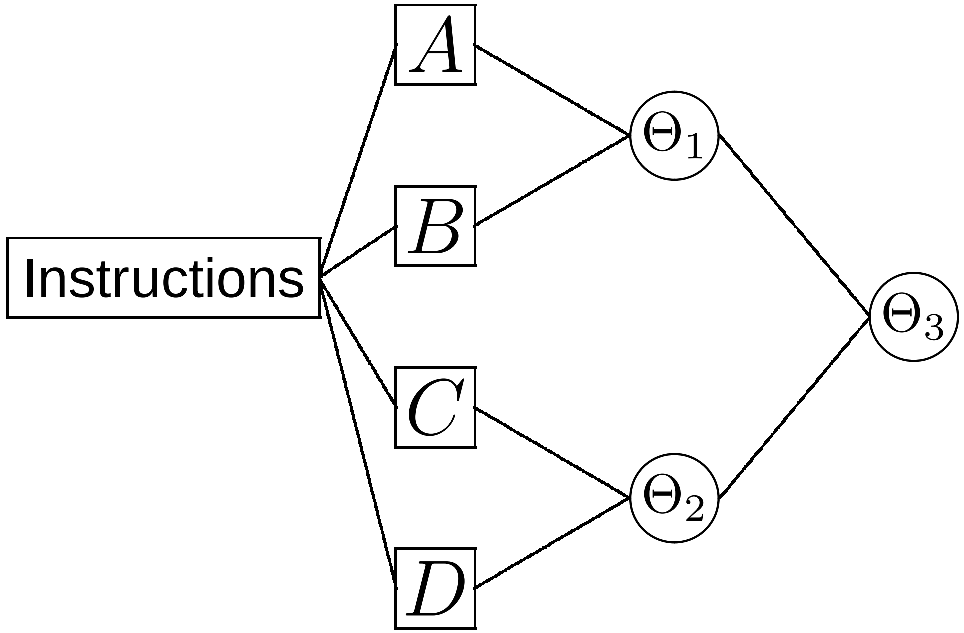
Extensions

Analysis of the ANES Data

Conclusion & Beyond

- Conclusion & Beyond
- Higher-Level Factors

This model did not fit the ANES data



i.e., $\phi_{11} = \ell_1^2$, $\phi_{12} = \ell_1 \ell_2$, and $\phi_{22} = \ell_2^2$.