Slide 1

Structural Equation Modeling with Ordinal Variables

Karl G Jöreskog, Uppsala University[a]

[a]email: karl.joreskog@dis.uu.se

Slide 2

Bartholomew (1985) and Bartholomew & Knott (1999) use the term latent variable models in a general setting. There are two sets of variables, manifest (observed) variables $\mathbf{x}$ and latent (unobserved) variables $\boldsymbol{\xi}$ with a joint distribution. If the marginal distribution $h(\boldsymbol{\xi})$ of $\boldsymbol{\xi}$ and the conditional distribution $g(\mathbf{x} \mid \boldsymbol{\xi})$ of $\mathbf{x}$ for given $\boldsymbol{\xi}$ exist, then the marginal distribution $f(\mathbf{x})$ of $\mathbf{x}$ must be

$$f(\mathbf{x}) = \int h(\boldsymbol{\xi}) g(\mathbf{x} \mid \boldsymbol{\xi}) d\boldsymbol{\xi} \ . \tag{1}$$

This is a tautology in the sense that it always holds if the distributions exist.

Karl G Jöreskog SAMSI Kickoff Paper

Slide 3

However, the idea of latent variable models is that the manifest variables should be independent for given latent variables, *i.e.*, the latent variables should account for all dependencies among the manifest variables. Thus,

$$g(\mathbf{x} \mid \boldsymbol{\xi}) = \prod_{i=1}^{p} g(x_i \mid \boldsymbol{\xi}) \ , \tag{2}$$

so that

$$f(\mathbf{x}) = \int h(\boldsymbol{\xi}) \prod_{i=1}^{p} g(x_i \mid \boldsymbol{\xi}) d\boldsymbol{\xi} \ . \tag{3}$$

To specify a latent variable model, one must therefore specify $h(\boldsymbol{\xi})$ and $g(x_i \mid \boldsymbol{\xi})$ for all i. The manifest variables may be continuous or categorical and the latent variables may be continuous or categorical. Thus there may be four classes of latent variable models:

Slide 4

Classification of Latent Variable Models

	Manifest Variables	
Latent Variables	Continuous	Categorical
Continuous	A: Factor Analysis Models	B: Latent Trait Models
Categorical	C: Latent Profile Models	D: Latent Class Models

Slide 5

Ordinal Variables

Observations on an ordinal variable represent responses to a set of ordered categories, such as a five-category Likert scale. It is only assumed that a person who selected one category has more of a characteristic than if he/she had chosen a lower category, but we do not know how much more. Ordinal variables are not continuous variables and should not be treated as if they are. It is common practice to treat scores 1, 2, 3, . . . assigned to categories as if they have metric properties but this is wrong. Ordinal variables do not have origins or units of measurements. Means, variances, and covariances of ordinal variables have no meaning. The only information we have are counts of cases in each cell of a multiway contingency table. To use ordinal variables in structural equation models requires other techniques than those that are traditionally employed with continuous variables.

Slide 6

- Models

  A Factor Models with Ordinal Indicators

  B Factor Models with Ordinal Indicators and Covariates

  C Factor Models with Ordinal and Continuous Indicators

  D Factor Models with Ordinal and Continuous Indicators and Covariates

  E General Structural Equation Models

- Data

  c Cross-sectional Data

  l Longitudinal Data

  m Multiple Group Data

- Methods

  BIV Bivariate Information Methods

  FUL Full Information Methods

Karl G Jöreskog SAMSI Kickoff Paper

Slide 7

Methods. Models and Data

Bivariate Information Methods

	Models				
Data	A	B	C	D	E
c	M	S	S	S	S
l	S	S	S	S	S
m	S	?	?	?	?

M=Much, S=Some, ?=Don't know

Slide 8

Full Information Methods

	Models				
Data	A	B	C	D	E
c	S	L	N	N	N
l	N	N	N	N	N
m	N	N	N	N	N

S=Some, L= Little, N=Nothing

Slide 9

Bivariate Information Methods

Model types A and B can be handled by using polychoric correlation or covariance matrices. This can be done for all three data types, see **ordinal.pdf** (81 pages) downloadable at `www.ssicentral.com/lisrel/ordinal.htm`.

Model types C, D, and E can be handled using polychoric and polyserial correlations and covariates.

These models and approaches are very easy to apply compared with full information methods. The only disadvantage seems to be that it is a three-stage procedure.

To avoid this I suggested (IMPS 2003) a BIML method in which all estimates are obtained by minimizing a single objective function. So far this has been developed for the exploratory factor analysis model (type A) but I am currently working on the extension to confirmatory factor analysis models with and without covariates (type B).

Slide 10

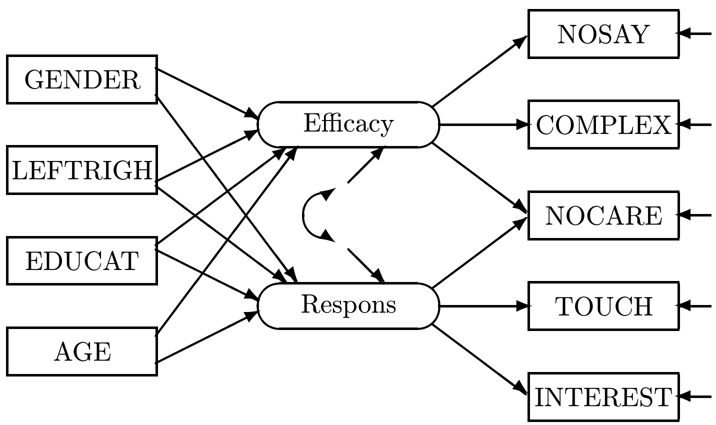
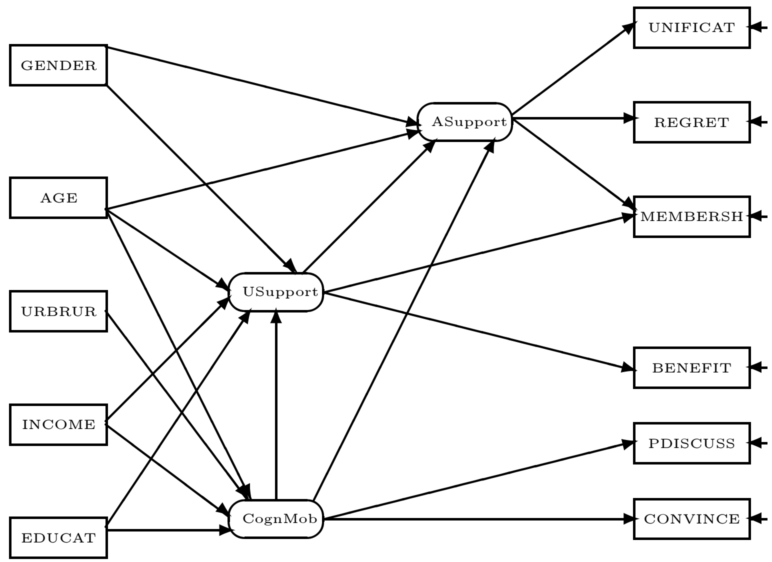


Figure 1: MIMIC Model for Efficacy and Respons

Karl G Jöreskog SAMSI Kickoff Paper

Slide 11



Slide 12

Full Information Methods

Most of the full information methods are limited to factor models (type A) There is a lot of literature on IRT models which usually assume a single latent variable. The multidimensional case have been discussed *e.g.*, by Moustaki & Knott (2000) and a comparison of these methods with the three-stage methods is presented Jöreskog & Moustaki (2001).

Moustaki (2003) extended these methods to allow for covariate effects on manifest and latent variable. This method is described in Moustaki, Jöreskog & Mavridis (2004) who compare them with the three-stage methods on the basis of two examples.

The biggest problem with the full information methods are associated with numerical integration. The integrals can be approximated by Gauss-Hermite quadrature, adaptive quadratic points, Monte Carlo methods, Laplace approximation, and Bayesian MCMC methods. There seems to be a lot of ongoing work on these methods.