Multilevel Analysis of Cross-National Surveys: The Role of Measurement and Estimation

Prof. Dr. Marco Steenbergen

University of Bern

February 11, 2011

Outline

1. Introduction
2. Measurement
3. Estimation
4. References

An Observation

- Much of the development of multilevel analysis (MLA) can be traced to educational research (e.g., Aitkin & Longford 1986; Raudenbush & Bryk 1986).
- This field has distinctive design and measurement features:
  - Large numbers of contextual units (e.g., schools) are sampled
  - Comparable measures of outcome variables (e.g., aptitude tests) exist

Another Observation

Cross-national surveys have quite different design and measurement features:

1. The comparability of survey measures is often questionable
2. The number of contextual units (i.e., countries) is usually limited and can rarely be considered a proper sample

An Implication

- It is tempting to use MLA for cross-national survey data.
- However, one should consider the distinctive design and measurement features of such data.
- This talk explains why and how.

Outline

1. Introduction
2. Measurement
3. Estimation
4. References

Preview

- Meaningful cross-national comparisons require measurement invariance.
- A lack of invariance can produce misleading conclusions in MLA.
- Specifically, invariance may produce "false" suggestions of random variation in intercepts and slopes.

A Measurement Model for Cross-National Surveys

Model

$$\mathbf{y}_{ij} = \boldsymbol{\tau}_j + \boldsymbol{\Lambda}_j \boldsymbol{\eta}_{ij} + \boldsymbol{\epsilon}_{ij}$$

Legend

- $j = 1 \cdots J$ is a particular country
- $i = 1 \cdots n_j$ is a particular respondent
- $\mathbf{y}_{ij}$ is a vector of p indicators
- $\boldsymbol{\eta}_{ij}$ is a vector of q latent variables
- $\boldsymbol{\epsilon}_{ij}$ is a $p \times 1$ vector of errors of measurement
- $\boldsymbol{\tau}_j$ is a $p \times 1$ vector of item intercepts
- $\boldsymbol{\Lambda}_j$ is a $p \times q$ matrix of factor loadings

Two Conditions for Measurement Invariance

1. Metric Invariance:
   - Scale metrics are identical across countries (Rock et al. 1978)
   - Thus, $\mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \cdots = \mathbf{\Lambda}_J$
2. Scalar Invariance:
   - There are no differential item biases across countries (Meredith 1993)
   - Thus, $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \cdots \boldsymbol{\tau}_J$

Lack of Invariance and MLA

- Imagine that intercept and slopes are constant across j at the latent variable level.
- In the absence of measurement invariance, the intercept and slopes at the manifest variable level will vary across j.

A Simple Example

Latent Variable Model

$$\eta_{ij} = \beta_0 + \beta_1 x_{ij} + \zeta_{ij}$$

Note that the intercept and slope are identical across j.

Measurement Model

$$y_{ij} = \tau_j + \lambda_j \eta_{ij} + \epsilon_{ij}$$

Note that the measure is not invariant across j.

A Simple Example

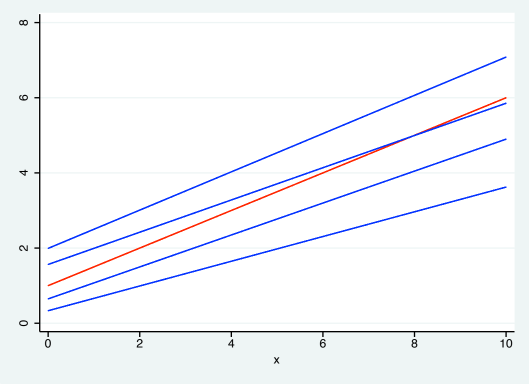
Combined Model

$$y_{ij} = \underbrace{\tau_j + \lambda_j \beta_0}_{Intercept} + \underbrace{\lambda_j \beta_1}_{Slope} x_{ij} + \underbrace{\lambda_j \zeta_{ij} + \epsilon_{ij}}_{Error}$$

Interpretation

Both the intercept and slope vary across countries.

A Simple Example



Red line depicts the relation between η and x, which is constant across all countries. Blue lines depict the relation between y and x for four different countries.

A Problem of Interpretation

- If an MLA shows random slopes/intercepts, this may have two different sources:
    1. The relationship between the underlying latent variable and a covariate varies across countries
    2. The measure of the latent variable is not invariant across countries
- Without further analysis we do not know which of these explains the results.

A Remedy

- The interpretational dilemma can be solved only when we can establish measurement invariance.
- This, however, requires multiple indicators.
- These are not always available.

Outline

1. Introduction
2. Measurement
3. Estimation
4. References

Preview

- Most studies suggest that proper MLA requires a large J.
- In many cross-national surveys, J is limited.
- Moreover it is often difficult to conceive of contexts as a probability sample, even though MLA assunes this.
- Estimation difficulties may be overcome using a Bayesian framework.

Sample Size Requirements

- Several studies:
  - Hox (1998), Kreft (1996), Mok (1995), Snijders & Bosker (1993)
- Clear consensus:
  - The bigger J, the better
  - With small J, variance component estimates can be severely biased (even with REML)
  - $J \geq 30$ according to Kreft (1996)
  - $J \geq 100$ according to Hox (1998)

Some Examples of J in Cross-National Surveys

- Commonly used surveys:
  - Eurobarometer: $J = 27$
  - European Social Survey (Round 4): $J = 31$
  - International Social Survey Programme (2008): $J = 47$
  - World Values Survey (2005-2008): $J = 58$
- Actual J may be lower due to:
  - Missing data
  - "Sample" restrictions (which may be necessary to satisfy exchangeability
- Thus, cross-national surveys often have dangerously low J.

What Is the Problem?

With small J,

- The log-likelihood does not contain much information about country-level variances, thus complicating estimation.
- Estimates of variance components are downwardly biased (e.g., Mok 1995).
- They may also be negative (Bian 2002).

An Additional Complication

- A 2-level design assumes that both individuals and contexts are sampled randomly from some population.
- In cross-national surveys the assumption of random sampling of individuals generally holds.
- However, the assumption of random sampling of countries is generally flawed:
    - Sometimes, the set of countries constitutes the population
    - At other times, it constitutes a non-probability sample

A Remedy

- Seltzer et al (1996) suggest that small-J problems may be overcome using a fully Bayesian approach.
- Such an approach also overcomes the sampling issue for countries.

How Bayesian Inference Helps

- Bayesian inference entails the specification of a prior, which adds information.
- The distinction between samples and populations is of secondary importance in Bayesian inference.
- Can be applied to linear and nonlinear models (cf. REML).

A Simple Example

Consider the following random effects ANOVA model:

$$
\begin{aligned}
y_{ij} &= \mu + \delta_j + \epsilon_{ij} \\
\delta_j &\sim N(0, \tau) \\
\epsilon_{ij} &\sim N(0, \sigma^2)
\end{aligned}
$$

A Simple Example

Assuming a balanced design, the kernel of the likelihood function is given by:

$$\mathcal{L}(D|\mu,\tau,\sigma^2) = \sigma^{-nJ}\left(1+\frac{n\tau}{\sigma^2}\right)^{-.5J} \times \exp\left[-\frac{1}{2\sigma^2}\sum_j (\mathbf{y}_j-\mu\boldsymbol{\iota})^T\left(\mathbf{I}-\frac{\tau}{\sigma^2+n\tau}\mathbf{J}\right)(\mathbf{y}_j-\mu\boldsymbol{\iota})\right]$$

where $\boldsymbol{\iota}$ is a vector of ones and $\mathbf{J}$ is a matrix of ones.

A Simple Example

- Bian (2002) proposes using independent inverted (conjugate) gamma priors.
- For example, the prior distribution for $\kappa = 1/\tau$ (the precision) would be:

$$p(\kappa) \quad \propto \quad \kappa^{\alpha+1} exp\left(-\beta\kappa\right)$$

- For small values of α and large values of β, this becomes uninformed.

A simple Example

- Bian (2002) performs a simulation study with $J = 3$, $n = 2$, $\mu = 5$, $\sigma^2 = 16$, and $tau = 4$.
- He conducts 1000 draws.
- The mean squared errors (MSEs) for τ display the following pattern:

$$MSE_{Bayes} < MSE_{FIML} < MSE_{REML}$$

Caveats

- Inserting strong priors that are wrong may make matters worse.
- Posteriors tend to be highly complex, even in the simple example from before:
    - Point estimates can generally not be obtained analytically
    - However, simulation-based methods (Gibbs sampler) are now readily available in software like MLwiN, R, and WinBUGS

Outline

1. Introduction
2. Measurement
3. Estimation
4. References

References

- Aitkin, Murray, and Nicholas Longford. 1986. Statistical Modelling Issues in School Effectiveness Studies. *Journal of the Royal Statistical Society, Series A (General)*, 149(1): 1-143.
- Bian, Guarui. 2002. Bayesian Estimates in a One-Way ANOVA Random Effects Model. *Australian & New Zealand Journal of Statistics* 44(1): 99-108.
- Hox, Joop. 1998. Multilevel Modeling: When and Why. In *Classification, Data Analysis, and Data Highways*, ed. Ingo Balderjahn, Rudolf Mathar, and Martin Schader. New York: Springer.
- Kreft, Ita G.G. 1996. Are Multilevel Techniques Necessary? An Overview, Including Simulation Studies. Los Angeles: California State University.
- Meredith, William. 1993. Measurement Invariance, Factor Analysis, and Factorial Invariance. *Psychometrika* 58(4) 525-543.
- Mok, Magdalena. 1995. Sample Size Requirements for 2-Level Designs in Educational Research. *Multilevel Modeling Newsletter* 7(2): 11-15.
- Raudenbush, Stephen, and Anthony S. Bryk. 1986. A Hierarchical Model for Studying School Effects. *Sociology of Education*, 59(1): 1-17.
- Rock, Donald A., Charles E. Werts, and Ronald L. Flaugher. 1978. The Use of Analysis of Covariance Structures for Comparing the Psychometric Properties of Multiple Variables Across Populations. *Multivariate Behavioral Research* 13(4): 403-418.
- Seltzer, Michael H., Wing H. Wong, and Anthony S. Bryk. 1996. Bayesian Analysis in Applications of Hierarchical Models: Issues and Methods. *Journal of Educational and Behavioral Statistics* 21(2): 131-167.
- Snijders, Tom A.B., and Roel J. Bosker. 1993. Standard Errors and Sample Sizes for Two-Level Research. *Journal of Educational Statistics* 18(3): 237-259.