Latent Variable Modeling Using Mplus: Day 3

Bengt Muthén & Tihomir Asparouhov

Mplus
www.statmodel.com

October, 2012

Table of Contents I

Table of Contents II

Table of Contents IV

1. Overview Of Day 3

More advanced day, focusing on the cutting-edge features in Version 7 related to multilevel analysis of complex survey data and item response theory (IRT) extensions.

Topics:

- IRT analysis, categorical factor analysis
  - Basic IRT
  - Intermediate IRT
- Multilevel analysis
  - Two-level analysis with random loadings (discriminations)
  - Three-level analysis
  - Cross-classified analysis
- Advanced IRT analysis
  - Group comparisons such as cross-national studies
  - Random items, G-theory
  - Random contexts
  - Longitudinal studies

Mplus Readings Related To Day 3

- Muthén (2008). Latent variable hybrids: Overview of old and new models. In Hancock, G. R., & Samuelsen, K. M. (Eds.), Advances in latent variable mixture models, pp. 1-24. Charlotte, NC: Information Age Publishing, Inc
- Asparouhov & Muthén (2012). Comparison of computational methods for high-dimensional item factor analysis. Technical Report. www.statmodel.com.
- Muthén & Asparouhov (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In J. Hox & J.K. Roberts (eds), Handbook of Advanced Multilevel Analysis, pp. 15-40. New York: Taylor and Francis
- Asparouhov & Muthén (2012). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. Technical Report. www.statmodel.com.

Let u_{ij} be a binary item j ($j = 1, 2, \dots p$) for individual i ($i = 1, 2, \dots n$), and express the probability of the outcome $u_{ij} = 1$ for this item as a function of m factors $\eta_{i1}, \eta_{i2}, \dots, \eta_{im}$ as follows,

$$P(u_{ij} = 1 \mid \eta_{i1}, \eta_{i2}, \dots, \eta_{im}) = F[-\tau_j + \sum_{k=1}^{m} \lambda_{jk}\, \eta_{ik}], \tag{1}$$

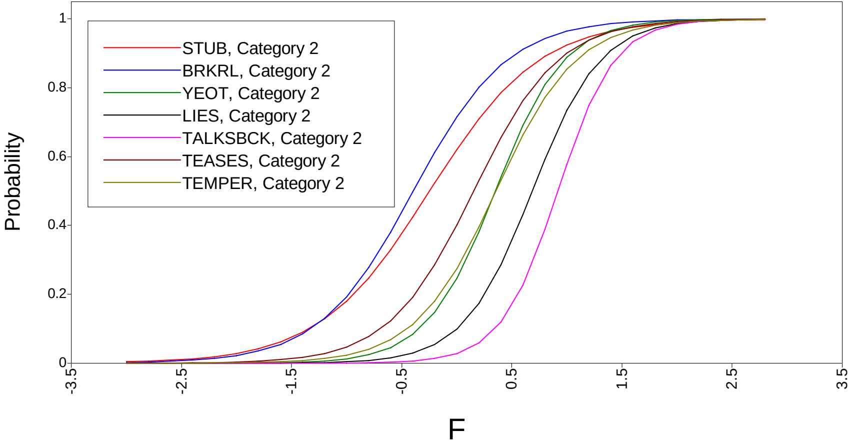
where with the logistic model and the general argument x, $F[x]$ represents the logistic function

$$F[x] = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}}, \tag{2}$$

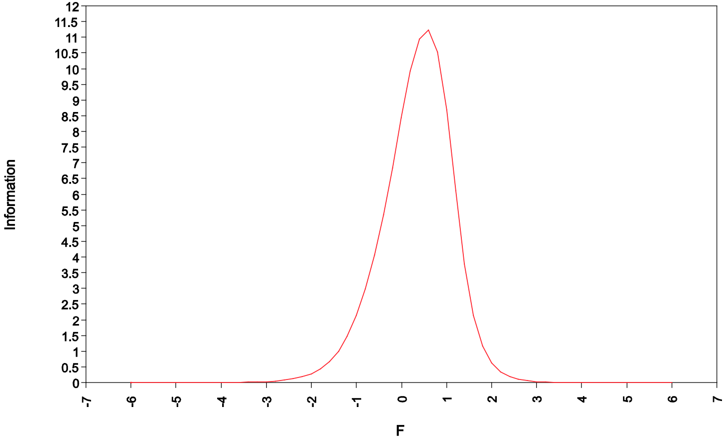
and with the probit model $F[x]$ represents the standard normal distribution function $\Phi[x]$.

The model is completed by assuming conditional independence among the items and normality for the factors.

Item Characteristic Curves From Maximum Likelihood IRT Analysis Of Seven Binary Aggression Items Measuring A Single Factor



Information Curve From Maximum Likelihood IRT Analysis Of Seven Binary Aggression Items Measuring A Single Factor



Mplus Offers Three Estimators For IRT And Factor Analysis Of Categorical Items

Criteria for comparison	Weighted least squares	Maximum likelihood	Bayes
Large number of factors	+	–	+
Large number of variables	–	+	+
Large number of subjects	+	–	–
Small number of subjects	–	+	+
Statistical efficiency	–	+	+
Missing data handling	–	+	+
Test of LRV structure	+	–	+
Ordered polytomous variables	+	–	–
Heywood cases	–	–	+
Zero cells	–	+	+
Residual correlations	+	–	±

Mplus Strengths For IRT And Categorical Factor Analysis

- High-dimensional analysis using WLSMV, Bayes, and ML two-tier
- Bi-factor EFA
- Modification indices, correlated residuals
- Multiple-group analysis
- Mixtures*
- Complex survey data handling: Stratification, weights
- Multilevel: two-level, three-level, and cross-classified
- Random loadings (discrimination) using Bayesian analysis
- Random item IRT
- Random subjects and contexts

* Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In Hancock, G. R., & Samuelsen, K. M. (Eds.), Advances in latent variable mixture models, pp. 1-24. Charlotte, NC: Information Age Publishing, Inc.

3. Bayesian EFA

- Bayesian estimation of exploratory factor analysis implemented in Mplus version 7 for models with continuous and categorical variables
- Asparouhov and Muthén (2012). Comparison of computational methods for high dimensional item factor analysis
- Asymptotically the Bayes EFA is the same as the ML solution
- Bayes EFA for categorical variable is a full information estimation method without using numerical integration and therefore feasible with any number of factors
- New in Mplus Version 7: Improved performance of ML-EFA for categorical variables, in particular high-dimensional EFA models with Montecarlo integration; improved unrotated starting values and standard errors

- The first step in the Bayesian estimation is the estimation of the unrotated model as a CFA model using the MCMC method
- Obtain posterior distribution for the unrotated solution
- To obtain the posterior distribution of the rotated parameters we simply rotate the generated unrotated parameters in every MCMC iteration, using oblique or orthogonal rotation
- No priors. Priors could be specified currently only for the unrotated solution
- If the unrotated estimation takes many iterations to converge, use THIN to reduce the number of rotations

Bayes EFA

- This MCMC estimation is complicated by identification issues that are similar to label switching in the Bayesian estimation of Mixture models
- There are two types of identification issues in the Bayes EFA estimation
- The first type is identification issues related to the unrotated parameters: loading sign switching
- Solution: constrain the sum of the loadings for each factor to be positive. Implemented in Mplus Version 7 for unrotated EFA and CFA. New in Mplus Version 7, leads to improved convergence in Bayesian SEM estimation

$$\sum_{i=1}^{p} \lambda_{ij} > 0$$

- The second type is identification issues related to the rotated parameters: loading sign switching and order of factor switching
- Solution: Align the signs s_j and factor order σ to minimize MSE between the current estimates λ and the average estimate from the previous MCMC iterations L

$$\sum_{i,j}(s_j\lambda_{i\sigma(j)} - L_{ij})^2$$

- Minimize over all sign allocations s_j and factor permutations σ

Bayes EFA

- Factor scores for the rotated solutions also available. Confidence intervals and posterior distribution plots
- Using the optimal rotation in each MCMC iteration we rotate the unrotated factors to obtain the posterior distribution of the rotated factors
- With continuous variables Bayes factor is computed to compare EFA with different number of factors. PPP value is computed with continuous or categorical variables

Bayes Factors

- Bayes factors is an easy and quick way to compare models using BIC

$$BF = \frac{P(H1)}{P(H0)} = \frac{Exp(-0.5BIC_{H1})}{Exp(-0.5BIC_{H0})}$$

- Values of BF greater than 3 are considered evidence in support of $H1$
- New in Mplus Version 7: BIC is now included for all models with continuous items (single level and no mixtures)
- The above method can be used to easily compare nested and non-nested models

Absolute bias, coverage and log-likelihood for EFA model with 7 factors and 35 ordered polytomous variables.

Method	λ_{11}	λ_{12}	Log-Likelihood
Mplus Monte 500	.01(0.97)	.00(0.83)	-28580.3
Mplus Monte 5000	.01(0.96)	.00(0.87)	-28578.4
Mplus Bayes	.01(.90)	.00(.96)	-
Mplus WLSMV	.00(.94)	.00(.89)	-
IRTPRO MHRM	.00(.54)	.00(.65)	-28665.2

Bayes EFA: Simulation Study ($n = 500$), Continued

Average standard error, ratio between average standard error and standard deviation for the EFA model with with 7 factors and ordered polytomous variables.

Method	λ_{11}	λ_{12}
Mplus Monte 500	0.033(1.00)	0.031(0.72)
Mplus Monte 5000	0.033(0.99)	0.035(0.81)
Mplus Bayes	0.030(0.97)	0.032(0.98)
Mplus WLSMV	0.030(0.97)	0.038(0.85)
IRTPRO MHRM	0.012(0.42)	0.026(0.65)

Bayes EFA is the most accurate full information estimation method for high-dimensional EFA with categorical variables.

Example is based on Mplus User's Guide example 4.1 generated with 4 factors and 12 indicators.

```
DATA: FILE IS ex4.1.dat;
VARIABLE: NAMES ARE y1-y12;
ANALYSIS: TYPE = EFA 1 5; estimator=bayes;
```

We estimate EFA with 1, 2, 3, 4 or 5 factors.

Bayes EFA: Results

Bayes factor results: The posterior probability that the number of factors is 4 is: 99.59%. However, this is a power result - there is enough information in the data to support 4 factors and not enough to support 5 factors. Use BITER = (10000)

```
POSTERIOR PROBABILITIES FOR ALL MODELS:
1-FACTOR MODEL   0.0000
2-FACTOR MODEL   0.0000
3-FACTOR MODEL   0.0041
4-FACTOR MODEL   0.9959
5-FACTOR MODEL   0.0000
```

Bayes EFA: Results

```
EXPLORATORY FACTOR ANALYSIS WITH 4 FACTOR(S):


Number of Free Parameters                              66

Bayesian Posterior Predictive Checking using Chi-Square

          95% Confidence Interval for the Difference Between
          the Observed and the Replicated Chi-Square Values

                        -35.978            41.864

          Posterior Predictive P-Value        0.376

Information Criterion

          Deviance (DIC)                               3231.767
          Estimated number of parameters (pD)            67.580
          Bayesian (BIC)                               3423.991
          Bayes factor for the model with 4 factor(s) against
          the model with 3 factor(s):
                                                       243.6367
          Log of Bayes factor for the model with 4 factor(s)
          against the model with 3 factor(s):
                                                         5.4957
```

Bayes EFA: Results

GEOMIN ROTATED LOADINGS (* significant at 5% level)

	1	2	3	4
Y1	0.622*	-0.008	0.028	-0.026
Y2	0.762*	-0.025	0.035	0.010
Y3	0.734*	0.055	-0.038	0.021
Y4	-0.112	0.592*	-0.008	-0.150
Y5	0.021	0.844*	0.049	0.007
Y6	0.053	0.639*	-0.029	-0.007
Y7	-0.009	0.037	0.788*	0.004
Y8	0.032	-0.059	0.646*	-0.115
Y9	0.008	0.040	0.602*	0.095
Y10	-0.042	0.065	0.005	0.671*
Y11	0.013	-0.078	-0.027	0.714*
Y12	0.051	-0.054	0.009	0.710*

4. Bayes Factor Scores Handling

- Version 7 uses improved language for factor scores with Bayesian estimation. The same language as for other estimators
- SAVEDATA: FILE=fs.dat; SAVE=FS(300); FACTORS=factor names; This command specifies that 300 imputations will be used to estimate the factor scores and that plausible value distributions are available for plotting
- Posterior mean, median, confidence intervals, standard error, all imputed values, distribution plot for each factor score for each latent variable for any model estimated with the Bayes estimator
- Bayes factor score advantages: more accurate than ML factor scores in small sample size, Bayes factor score more accurate in secondary analysis such as for example computing correlations between factor

Bayes Factor Scores Example

- Asparouhov & Muthén (2010). Plausible values for latent variables using Mplus
- Factor analysis with 3 indicators and 1 factor. Simulated data with N=45. True factor values are known. Bayes factor score estimates are more accurate. Bayes factor score SE are more accurate
- ML factor scores are particularly unreliable when Var(Y) is near 0

	ML	Bayes
MSE	0.636	0.563
Coverage	20%	89%
Average SE	0.109	0.484

5. Two-Level Analysis With Random Intercepts (Difficulties) And Random Loadings (Discrimination)

- Measurement invariance across groups
- Overview and an example of hospital ratings (continuous items)
- Two-level random loadings in IRT using the PISA math data (binary items)
- Testing for non-zero variance of random loadings
- Individual differences factor analysis

5.1 Advances In Multiple-Group Analysis: Invariance Across Groups

- An old dilemma
- Two new solutions

Fixed Versus Random Groups

- Fixed mode:
  - Inference to only the groups in the sample
  - Small to medium number of groups
- Random mode:
  - Inference to a population of groups from which the current set of groups is a random sample
  - Medium to large number of groups

Two New Solutions In Mplus Version 7

- New solution no. 1, suitable for a small to medium number of groups
  - A new BSEM approach where group is a fixed mode: Multiple-group BSEM (see Utrecht video, Part 1 handout)
  - Approximate invariance allowed
- New solution no. 2, suitable for a medium to large number of groups
  - A new Bayes approach where group is a random mode
  - No limit on the number of groups

5.1.1 Hospital Data Example

Shortell et al. (1995). Assessing the impact of continuous quality improvement/total quality management: concept versus implementation. Health Services Research, 30, 377-401.

- Survey of 67 hospitals, $n = 7168$ employee respondents, approximately 100/hospital
- 6 dimensions of an overall ”quality improvement implementation” based on the Malcom Baldrige National Quality Award critera
- Focus on 10 items measuring a leadership dimension
- Continuous items

Hospital Data: Old And New Factor Analysis Alternatives

- Hospital as Fixed Mode:
  - Old approach: Conventional multiple-group factor analysis
  - New approach: BSEM multiple-group factor analysis
- Hospital as Random Mode:
  - Old approach: Conventional two-level factor analysis
  - New approach: Bayes random loadings two-level factor analysis (random factor variances also possible)

5.1.2 Hospital As Fixed Mode: Conventional Multiple-Group Factor Analysis

Regular ML analysis:

```
VARIABLE:    USEVARIABLES = lead21-lead30! info31-info37
             ! straqp38-straqp44 hru45-hru52 qm53-qm58 hosp;
             MISSING = ALL(-999);
             !CLUSTER = hosp;
             GROUPING = hosp (101 102 104 105 201 301-306
             308 310-314 316-320 322 401-403 405-409 412-416
             501-503 505-512 602-609 612-613 701 801 901-908);
ANALYSIS:    ESTIMATOR = ML;
             PROCESSORS = 8;
MODEL:
             lead BY lead21-lead30; ! specifies measurement invariance
PLOT:        TYPE = PLOT2;
OUTPUT:      TECH1 TECH8 MODINDICES(ALL);
```

Hospital As Fixed Mode: Conventional Multiple-Group Factor Analysis, Continued

Maximum-likelihood analysis with χ^2 test of model fit and modification indices.

Holding measurement parameters equal across groups/hospitals results in poor fit with many moderate-sized modification indices and none that sticks out as much larger than the others.

Conventional multiple-group factor analysis ”fails”.

5.1.3 Group As Random Mode: Conventional Two-Level Factor Analysis

- Recall random effects ANOVA (individual i in cluster j):

$$y_{ij} = \nu + \eta_j + \varepsilon_{ij} = y_{B_j} + y_{W_j} \tag{3}$$

- Two-level factor analysis ($r = 1, 2, \ldots, p$ items; 1 factor on each level):

$$y_{rij} = \nu_r + \lambda_{B_r}\, \eta_{B_j} + \varepsilon_{B_{rj}} + \lambda_{W_{ij}}\, \eta_{W_{ij}} + \varepsilon_{W_{rij}} \tag{4}$$

- Alternative expression often used in 2-level IRT:

$$y_{rij} = \nu_r + \lambda_r\, \eta_{ij} + \varepsilon_{rij}, \tag{5}$$

$$\eta_{ij} = \eta_{B_j} + \eta_{W_{ij}}, \tag{6}$$

so that λ is the same for between and within.

Input Excerpts For Hospital As Random Mode: Conventional Two-Level Factor Analysis

```
               USEVARIABLES = lead21-lead30;
               MISSING = ALL (-999);
               CLUSTER = hosp;
ANALYSIS:      TYPE = TWOLEVEL;
               ESTIMATOR = ML;
               PROCESSORS = 8;
MODEL:         %WITHIN%
               leadw BY lead21-lead30* (lam1-lam10);
               leadw@1;
               %BETWEEN%
               leadb BY lead21-lead30* (lam1-lam10);
               leadb;
OUTPUT:        TECH1 TECH8 MODINDICES(ALL);
```

Results For Hospital As Random Mode: Conventional Two-Level Factor Analysis

Equality of within- and between-level factor loadings cannot be rejected by χ^2 difference testing

10 % of the total variance in the leadership factor is due to between-hospital variation

No information about measurement invariance across hospitals

5.2 New Solution No. 2: Group Is Random Mode Two-Level Factor Analysis With Random Loadings

Consider a single factor η. For factor indicator r ($r = 1, 2, \ldots p$) for individual i in group (cluster) j,

$$y_{rij} = \nu_{rj} + \lambda_{rj}\,\eta_{ij} + \varepsilon_{ij}, \tag{7}$$

$$\eta_{ij} = \eta_j + \zeta_{ij}, (\textit{this may be viewed as } \eta_{B_j} + \eta_{W_{ij}}) \tag{8}$$

$$\nu_{rj} = \nu_r + \delta_{\nu_j}, \tag{9}$$

$$\lambda_{rj} = \lambda_r + \delta_{\lambda_j}, \tag{10}$$

where ν_r is the mean of the r^{th} intercept and λ_r is the mean of the r^{th} factor loading. Because the factor loadings are free, the factor metric is set by fixing $V(\zeta_{ij}) = 1$ (the between-level variance $V(\eta_j)$ is free). Note that the same loading is multiplying both the between- and within-level parts of the factor η.

Two-Level Factor Analysis With Random Loadings: 3 Model Versions

$$y_{rij} = \nu_{rj} + \lambda_{rj}\, \eta_{ij} + \varepsilon_{ij}, \tag{11}$$

$$\eta_{ij} = \eta_j + \zeta_{ij}, (\textit{this may be viewed as } \eta_{B_j} + \eta_{W_{ij}}) \tag{12}$$

$$\nu_{rj} = \nu_r + \delta_{\nu_j}, \tag{13}$$

$$\lambda_{rj} = \lambda_r + \delta_{\lambda_j}, \tag{14}$$

A first alternative to this model is that $V(\eta_j) = 0$ so that the factor with random loadings has only within-level variation. Instead, there can be a separate between-level factor with non-random loadings, measured by the random intercepts of the y indicators as in regular two-level factor analysis, $y_{rj} = \lambda_{B_r}\, \eta_{B_j} + \zeta_{rj}$, where y_{rj} is the between part of y_{rij}.

A second alternative is that the λ_{B_r} loadings are equal to the means of the random loadings λ_r.

Part 1: Random factor loadings (decomposition of the factor into within- and between-level parts)

```
TITLE:     this is an example of a two-level MIMIC
           model with continuous factor indicators,
           random factor loadings, two covariates on
           within, and one covariate on between
           with equal loadings across levels
DATA:      FILE = ex9.19.dat;
VARIABLE:  NAMES = y1-y4 x1 x2 w clus;
           WITHIN = x1 x2;
           BETWEEN = w;
           CLUSTER = clus;
ANALYSIS:  TYPE = TWOLEVEL RANDOM;
           ESTIMATOR = BAYES;
           PROCESSORS = 2;
           BITER = (1000);
MODEL:     %WITHIN%
           s1-s4 | f BY y1-y4;
           f@1;
           f ON x1 x2;
           %BETWEEN%
           f ON w;
           f; ! defaults: s1-s4; [s1-s4];
PLOT:      TYPE = PLOT2;
OUTPUT:    TECH1 TECH8;
```

New Solution No. 2: Group Is Random Mode. UG Ex9.19

Part 2: Random factor loadings and a separate between-level factor

```
MODEL:     %WITHIN%
           s1-s4 | f BY y1-y4;
           f@1;
           f ON x1 x2;
           %BETWEEN%
           fb BY y1-y4;
           fb ON w;
```

f@0; is the between-level default

Part 3: Random factor loadings and a separate between-level factor with loadings equal to the mean of the random loadings

```
MODEL:     %WITHIN%
           s1-s4 | f BY y1-y4;
           f@1;
           f ON x1 x2;
           %BETWEEN%
           fb BY y1-y4* (lam1-lam4);
           fb ON w;
           [s1-s4*1] (lam1-lam4);
```

5.2.2 Monte Carlo Simulations For Groups As Random Mode: Two-Level Random Loadings Modeling

- The effect of treating random loadings as fixed parameters
  - Continuous variables
  - Categorical variables
- Small number of clusters/groups

The Effect Of Treating Random Loadings As Fixed Parameters With Continuous Variables

Table: Absolute bias and coverage for factor analysis model with random loadings - comparing random intercepts and loadings and v.s. random intercepts and fixed loadings models

parameter	Bayes	ML with fixed loadings
θ_1	0.00(0.97)	0.20(0.23)
μ_1	0.01(0.95)	0.14(0.66)
λ_1	0.01(0.96)	0.00(0.80)
θ_2	0.02(0.89)	0.00(0.93)

Ignoring the random loadings leads to biased mean and variance parameters and poor coverage. The loading is unbiased but has poor coverage.

The Effect Of Treating Random Loadings As Fixed Parameters With Categorical Variables

Table: Absolute bias and coverage for factor analysis model with categorical data and random loadings - comparing random loadings and intercepts v.s. random intercepts and fixed loadings models

parameter	Bayes	WLSMV with fixed loadings
τ_1	0.05(0.96)	0.17(0.63)
λ_1	0.03(0.92)	0.13(0.39)
θ_2	0.05(0.91)	0.11(0.70)

Ignoring the random loadings leads to biased mean, loading and variance parameters and poor coverage.

Random Loadings With Small Number Of Clusters/Groups

- Many applications have small number of clusters/groups. How many variables and random effects can we use?
- Independent random effects model - works well even with 50 variables (100 random effects) and 10 clusters
- Weakly informative priors are needed to eliminate biases for cluster level variance parameters
- Correlated random effects model (1-factor model) - works only when "number of clusters $>$ number of random effects". More than 10 clusters are needed with 5 variables or more.
- What happens if you ignore the correlation: standard error underestimation, decreased accuracy in cluster specific estimates
- BSEM: Muthén, B. and Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. Forthcoming in Psychological Methods.
- Using BSEM with 1-factor model for the random effects and tiny priors $N(1,\sigma)$ for the loadings resolves the problem.

5.3 Two-Level Random Loadings In IRT: The PISA Data

- Fox, J.-P., and A. J. Verhagen (2011). Random item effects modeling for cross-national survey data. In E. Davidov & P. Schmidt, and J. Billiet (Eds.), Cross-cultural Analysis: Methods and Applications
- Fox (2010). Bayesian Item Response Modeling. Springer
- Program for International Student Assessment (PISA 2003)
- 9,769 students across 40 countries
- 8 binary math items

Random Loadings In IRT

- Y_{ijk} - outcome for student i, in country j and item k

$$P(Y_{ijk} = 1) = \Phi(a_{jk}\theta_{ij} + b_{jk})$$

$$a_{jk} \sim N(a_k, \sigma_{a,k}), b_{jk} \sim N(b_k, \sigma_{b,k})$$

  Both discrimination (a) and difficulty (b) vary across country
- The θ ability factor is decomposed as

$$\theta_{ij} = \theta_j + \varepsilon_{ij}$$

$$\theta_j \sim N(0, v), \varepsilon_{ij} \sim N(0, v_j), \sqrt{v_j} \sim N(1, \sigma)$$

- The mean and variance of the ability vary across country
- For identification purposes the mean of $\sqrt{v_j}$ is fixed to 1, this replaces the traditional identification condition that $v_j = 1$
- Model preserves common measurement scale while accommodating measurement non-invariance as long as the variation in the loadings is not big

Random Loadings In IRT, Outline

- Three two-level factor models with random loadings
- Testing for significance of the random loadings
- Two methods for adding cluster specific factor variance in addition to the random loadings
- All models can be used with continuous outcomes as well

Random Loadings In IRT Continued

- Model 1 - without cluster specific factor variance, cluster specific discrimination, cluster specific difficulty, cluster specific factor mean

$$P(Y_{ijk} = 1) = \Phi(a_{jk}\theta_{ij} + b_{jk})$$

$$a_{jk} \sim N(a_k, \sigma_{a,k}), b_{jk} \sim N(b_k, \sigma_{b,k})$$

$$\theta_{ij} = \theta_j + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, 1)$$

$$\theta_j \sim N(0, v)$$

Random Loadings In IRT Continued

- Note that cluster specific factor variance is confounded with cluster specific factor loadings (it is not straight forward to separate the two). Ignoring cluster specific factor variance should not lead to misfit. It just increases variation in the factor loadings which absorbs the variation in the factor variance
- Model 1 setup in Mplus: the factor f is used on both levels to represent the within ε_{ij} and the between θ_j part of the factor

```
model:
            %within%
            s1-s8 | f by y1-y8; f@1;

            %between%
            f y1-y8 s1-s8;
```

- All between level components are estimated as independent. Dependence can be introduced by adding factor models on the between level or covariances

PISA Results - Discrimination (Mean Of Random Loadings) And Difficulty

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	95% C.I. Upper 2.5%
Between Level					
Means					
S1	0.735	0.037	0.000	0.666	0.820
S2	1.036	0.058	0.000	0.931	1.158
S3	0.631	0.026	0.000	0.588	0.687
S4	0.622	0.031	0.000	0.558	0.678
S5	0.528	0.032	0.000	0.466	0.593
S6	0.353	0.035	0.000	0.287	0.421
S7	0.607	0.032	0.000	0.549	0.677
S8	0.616	0.034	0.000	0.561	0.686
Thresholds					
Y1$1	-0.481	0.055	0.000	-0.581	-0.372
Y2$1	0.371	0.072	0.000	0.238	0.519
Y3$1	0.059	0.047	0.115	-0.037	0.145
Y4$1	-0.270	0.049	0.000	-0.368	-0.172
Y5$1	0.059	0.037	0.056	-0.013	0.125
Y6$1	-1.496	0.044	0.000	-1.584	-1.409
Y7$1	-0.691	0.045	0.000	-0.777	-0.605
Y8$1	-0.862	0.038	0.000	-0.926	-0.786

PISA Results - Random Variation Across Countries

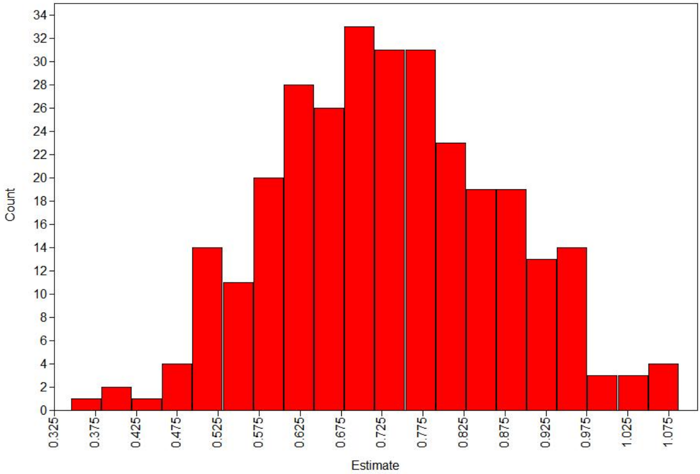
	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%
Between Level					
Variances					
Y1	0.044	0.017	0.000	0.022	0.087
Y2	0.063	0.025	0.000	0.030	0.124
Y3	0.029	0.011	0.000	0.013	0.057
Y4	0.034	0.012	0.000	0.018	0.066
Y5	0.011	0.006	0.000	0.003	0.027
Y6	0.034	0.016	0.000	0.013	0.075
Y7	0.023	0.011	0.000	0.008	0.050
Y8	0.007	0.006	0.000	0.001	0.023
F	0.317	0.085	0.000	0.204	0.528
S1	0.013	0.010	0.000	0.002	0.037
S2	0.078	0.040	0.000	0.027	0.181
S3	0.006	0.005	0.000	0.001	0.022
S4	0.012	0.008	0.000	0.002	0.031
S5	0.017	0.010	0.000	0.005	0.044
S6	0.020	0.012	0.000	0.005	0.048
S7	0.006	0.007	0.000	0.001	0.027
S8	0.011	0.008	0.000	0.002	0.033

Country-Specific Mean Ability Parameter

Factor scores can be obtained for the mean ability parameter using the country specific factor loadings. Highest and lowest 3 countries.

Country	Estimate and confidence limits
FIN	0.749 (0.384 , 0.954)
KOR	0.672 (0.360 , 0.863)
MAC	0.616 (0.267 , 1.041)
BRA	-0.917 (-1.166 , -0.701)
IDN	-1.114 (-1.477 , -0.912)
TUN	-1.156 (-1.533 , -0.971)

Country-Specific Distribution For The Mean Ability Parameter For FIN



Random Loadings In IRT Continued

- Random loadings have small variances, however even small variance of 0.01 implies a range for the loading of 4*SD=0.4, i.e., substantial variation in the loadings across countries
- How can we test significance for the variance components? If variance is not near zero the confidence intervals are reliable. However, when the variance is near 0 the confidence interval does not provide evidence for statistical significance
- Example: Var(S2)=0.078 with confidence interval [0.027,0.181] is significant but Var(S7)=0.006 with confidence interval [0.001,0.027] is not clear. Caution: if the number of clusters on the between level is small all these estimates will be sensitive to the prior

5.4 Testing For Non-Zero Variance Of Random Loadings

- Verhagen & Fox (2012) Bayesian Tests of Measurement Invariance
- Test the null hypothesis $\sigma = 0$ using Bayesian methodology
- Substitute null hypothesis $\sigma < 0.001$
- Estimate the model with σ prior IG(1,0.005) with mode 0.0025 (If we push the variances to zero with the prior, would the data provide any resistance?)

$$BF = \frac{P(H_0)}{P(H_1)} = \frac{P(\sigma < 0.001|data)}{P(\sigma < 0.001)} = \frac{P(\sigma < 0.001|data)}{0.7\%}$$

- $BF > 3$ indicates loading has 0 variance, i.e., loading invariance

Testing For Non-Zero Variance Of Random Loadings

- Other cutoff values are possible such as 0.0001 or 0.01
- Implemented in Mplus in Tech16
- Estimation should be done in two steps. First estimate a model with non-informative priors. Second in a second run estimate the model with IG(1,0.005) variance prior to test the significance
- How well does this work? The problem of testing for zero variance components is difficult. ML T-test or LRT doesn't provide good solution because it is a borderline testing
- New method which is not studied well but there is no alternative particularly for the case of random loadings. The random loading model can not be estimated with ML due to too many dimensions of numerical integration

Testing For Non-Zero Variance Of Random Loadings

- Simulation: Simple factor analysis model with 5 indicators, N=2000, variance of factor is free, first loading fixed to 1. Simulate data with Var(f)=0.0000001. Using different BITER commands with different number of min iterations
- BITER=100000; rejects the non-zero variance hypothesis 51% of the time
- BITER=100000(5000); rejects the non-zero variance hypothesis 95% of the time
- BITER=100000(10000); rejects the non-zero variance hypothesis 100% of the time
- Conclusion: The variance component test needs good number of iterations due to estimation of tail probabilities
- Power: if we generate data with Var(f)=0.05, the power to detect significantly non-zero variance component is 50% comparable to ML T-test of 44%

Testing For Non-Zero Variance Of Random Loadings In The PISA Model

Add IG(1,0.005) prior for the variances we want to test

```
MODEL:
        %WITHIN%
        s1-s8 | f BY y1-y8;
        f@1;
        %BETWEEN%
        f;
        y1-y8 (v1-v8);
        s1-s8 (v9-v16);
MODEL PRIORS:
        v1-v16~IG(1, 0.005);
OUTPUT:
        TECH1 TECH16;
```

Testing For Non-Zero Variance Of Random Loadings In The PISA Model

- Bayes factor greater than 3 in any column indicate non-significance (at the corresponding level). For example, Bayes factor greater than 3 in the second column indicates variance is less than 0.001.
- Bayes factor=10 in column 3 means that a model with variance smaller than 0.001 is 10 times more likely than a model with non-zero variance
- The small variance prior that is used applies to a particular variance threshold hypothesis. For example, if you want to test the hypothesis $v < 0.001$, use the prior $v \sim$ IG(1,0.005), and look for the results in the second column. If you want to test the hypothesis $v < 0.01$, use the prior $v \sim$ IG(1,0.05), and look for the results in the third column.
- Parameters 9-16 variances of the difficulty parameters
- Parameters 26-33 variances of the discrimination parameters

Results: TECH16

Testing significance of variance components

Parameter	BF for <0.0001	BF for <0.001	BF for <0.01
Parameter 9	0.0000	0.0000	0.0000
Parameter 10	0.0000	0.0000	0.0000
Parameter 11	0.0000	0.0000	0.0115
Parameter 12	0.0000	0.0000	0.0041
Parameter 13	0.0000	0.0000	0.5431
Parameter 14	0.0000	0.0000	0.0324
Parameter 15	0.0000	0.0000	0.0417
Parameter 16	0.0000	0.7093	1.3432
Parameter 26	0.0000	2.4226	1.5264
Parameter 27	0.0000	0.0982	0.7791
Parameter 28	0.0000	4.2996	1.5951
Parameter 29	0.0000	1.0476	1.4318
Parameter 30	0.0000	0.4911	1.2564
Parameter 31	0.0000	0.3929	0.9429
Parameter 32	0.0000	2.4117	1.5428
Parameter 33	0.0000	1.1895	1.4523

Random Loadings In IRT Continued

Estimate a model with fixed and random loadings. Loading 3 is now a fixed parameter rather than random.

```
MODEL:
        %WITHIN%
        f@1;
        s1-s2 | f BY y1-y2;
        f BY y3*1;
        s4-s8 | f BY y4-y8;
        %BETWEEN%
        f;
        y1-y8;
        s1-s8;
```

Random Loadings In IRT Continued

- Model 2 - Between level factor has different (non-random) loadings

$$P(Y_{ijk} = 1) = \Phi(a_{jk}\theta_{ij} + c_k\theta_j + b_{jk})$$

$$a_{jk} \sim N(a_k, \sigma_{a,k}), b_{jk} \sim N(b_k, \sigma_{b,k})$$

$$\theta_{ij} \sim N(0,1)$$

$$\theta_j \sim N(0,1)$$

- Model 2 doesn't have the interpretation that θ_j is the between part of the θ_{ij} since the loadings are different

Random Loadings In IRT Continued

- Model 3 - Between level factor has loadings equal to the mean of the random loadings

$$P(Y_{ijk} = 1) = \Phi(a_{jk}\theta_{ij} + a_k\theta_j + b_{jk})$$

$$a_{jk} \sim N(a_k, \sigma_{a,k}), b_{jk} \sim N(b_k, \sigma_{b,k})$$

$$\theta_{ij} \sim N(0,1)$$

$$\theta_j \sim N(0,v)$$

- Model 3 has the interpretation that θ_j is approximately the between part of the θ_{ij}
- Model 3 is nested within Model 2 and can be tested by testing the proportionality of between and within loadings

Random Loadings In IRT Continued

Model 3 setup. The within factor f now represents only θ_{ij}, fb represents θ_j.

```
model:
            %within%
            s1-s8 | f by y1-y8;
            f@1;

            %between%
            y1-y8 s1-s8;
            [s1-s8*1] (p1-p8);
            fb by y1-y8*1 (p1-p8);
            fb;
```

Random Loadings In IRT Continued: Adding Cluster Specific Factor Variance: Method 1

Replace $Var(\theta_{ij}) = 1$ with $Var(\theta_{ij}) = 0.51 + (0.7 + \sigma_j)^2$ where σ_j is a zero mean cluster level random effect. The constant 0.51 is needed to avoid variances fixed to 0 which cause poor mixing. This approach can be used for any variance component on the within level.

```
model:
          %within%
          s1-s8 | f by y1-y8;
          sigma | e by f;
          e@1;
          f@0.51;

          %between%
          y1-y8 s1-s8;
          [s1-s8*1] (p1-p8);
          fb by y1-y8*1 (p1-p8);
          fb sigma;
          [sigma@0.7];
```

Random Loadings In IRT Continued: Adding Cluster Specific Factor Variance: Method 2

- Variability in the loadings is confounded with variability in the factor variance
- A model is needed that can naturally separate the across-country variation in the factor loadings and the across-country variation in the factor variance
- From a practical perspective we want to have as much variation in the factor variance and as little as possible in the factor loadings to pursue the concept of measurement invariance or approximate measurement invariance

Random Loadings In IRT Continued: Adding Cluster Specific Factor Variance: Method 2, Cont'd

- Replace $Var(\theta_{ij}) = 1$ with $Var(\theta_{ij}) = (1+\sigma_j)^2$ where σ_j is a zero mean cluster level random effect. This model is equivalent to having $Var(\theta_{ij}) = 1$ and the discrimination parameters as

$$a_{jk} = (1+\sigma_j)(a_k+\varepsilon_{jk})$$

- Because σ_j and ε_{jk} are generally small, the product $\sigma_j \cdot \varepsilon_{jk}$ is of smaller magnitude so it is ignored
-

$$a_{jk} \approx a_k + \varepsilon_{jk} + a_k \sigma_j$$

- σ_j can be interpreted as **between level latent factor for the random loadings** with loadings a_k equal to the means of the random loadings

Random Loadings In IRT Continued: Adding Cluster Specific Factor Variance: Method 2, Cont'd

- Factor analysis estimation tends to absorb most of the correlation between the indicators within the factor model and to minimize the residual variances
- Thus the model will try to explain as much as possible the variation between the correlation matrices across individual as a variation in the factor variance rather than as a variation in the factor loadings.
- Thus this model is ideal for evaluating and separating the loading non-invariance and the factor variance non-invariance
- Testing $Var(\varepsilon_{jk}) = 0$ is essentially a test for measurement invariance. Testing $Var(\sigma_j) = 0$ is essentially a test for factor variance invariance across the cluster

Random Loadings In IRT Continued: Adding Cluster Specific Factor Variance: Method 2

Method 2 setup. Optimal in terms of mixing and convergence.

MODEL:

```
%WITHIN%
s1-s8 | f BY y1-y8;
f@1;
%BETWEEN%
y1-y8 s1-s8;
[s1-s8*1] (p1-p8);
fb BY y1-y8*1 (p1-p8);
sigma BY s1-s8*1 (p1-p8);
fb sigma;
```

Asparouhov & Muthén (2012). General Random Effect Latent Variable Modeling: Random Subjects, Items, Contexts, and Parameters.

5.5 Two-Level Random Loadings: Individual Differences Factor Analysis

- Jahng S., Wood, P. K.,& Trull, T. J., (2008). Analysis of Affective Instability in Ecological Momentary Assessment: Indices Using Successive Difference and Group Comparison via Multilevel Modeling. Psychological Methods, 13, 354-375
- An example of the growing amount of EMA data
- 84 borderline personality disorder (BPD) patients. The mood factor for each individual is measured with 21 self-rated continuous items. Each individual is measured several times a day for 4 weeks for total of about 100 assessments
- Factor analysis is done as a two-level model where cluster=individual, many assessments per cluster

Individual Differences Factor Analysis

- This data set is perfect to check if a measurement instrument is interpreted the same way by different individuals. Some individuals response may be more correlated for some items, i.e., the factor analysis should be different for different individuals.
- Example: suppose that one individual answers item 1 and 2 always the same way and a second individual doesn't. We need separate factor analysis models for the two individuals, individually specific factor loadings.
- If the within level correlation matrix varies across cluster that means that the loadings are individually specific
- Should in general factors loadings be individually specific? This analysis can NOT be done in cross-sectional studies, only longitudinal studies with multiple assessments

- Large across-time variance of the mood factor is considered a core feature of BPD that distinguishes this disorder from other disorders like depressive disorders.
- The individual-specific factor variance is the most important feature in this study
- The individual-specific factor variance is confounded with individual-specific factor loadings
- How to separate the two? Answer: **Factor Model for the Random Factor Loadings** as in the PISA data

Let Y_{pij} be item p, for individual i, at assessment j. Let X_i be an individual covariate. The model is given by

$$Y_{pij} = \mu_p + \zeta_{pi} + s_{pi}\eta_{ij} + \varepsilon_{pij}$$

$$\eta_{ij} = \eta_i + \beta_1 X_i + \xi_{ij}$$

$$s_{pi} = \lambda_p + \lambda_p \sigma_i + \varepsilon_{pi}$$

$$\sigma_i = \beta_2 X_i + \zeta_i$$

β_1 and β_2 represent the effect of the covariate X on the mean and the variance of the mood factor.

IDFA has individually specific: item mean, item loading, factor mean, factor variance.

Individual Differences Factor Analysis Model Setup

Many different ways to set up this model in Mplus. The setup below gives the best mixing/convergence performance.

```
MODEL:
        %WITHIN%
        s1-s21 | f BY jittery-scornful;
        f@1;
        %BETWEEN%
        f ON x; f;
        s1-s21 jittery-scornful;
        [s1-s21*1] (lambda1-lambda21);
        sigma BY s1-s21*1 (lambda1-lambda21);
        sigma ON x; sigma;
```

Individual Differences Factor Analysis Results

All variance components are significant. Percent Loading Invariance = the percentage of the variation of the loadings that is explained by factor variance variation.

item	Res Var	Mean	Var of Mean	Loading	Var of Loading	Percent Loading Invariance
Item 1	0.444	1.505	0.287	0.261	0.045	0.29
Item 2	0.628	1.524	0.482	0.377	0.080	0.32
Item 3	0.331	1.209	0.057	0.556	0.025	0.77
Item 4	0.343	1.301	0.097	0.553	0.030	0.73
Item 5	0.304	1.094	0.017	0.483	0.053	0.54

Individual Differences Factor Analysis Conclusions

- Clear evidence that measurement items are not interpreted the same way by different individuals and thus individual-specific adjustments are needed to the measurement model to properly evaluate the underlying factors: IDFA model
- IDFA model clearly separates factor variance variation from the factor loadings variation
- Asparouhov & Muthén, B. (2012). General Random Effect Latent Variable Modeling: Random Subjects, Items, Contexts, and Parameters

6. 3-Level Analysis

Continuous outcomes: ML and Bayesian estimation

Categorical outcomes: Bayesian estimation (Bayes uses probit)

Count and nominal outcomes: Not yet available

6.1 Types Of Observed Variables In 3-Level Analysis

Each Y variable is decomposed as

$$Y_{ijk} = Y_{1ijk} + Y_{2jk} + Y_{3k},$$

where Y_{1ijk}, Y_{2jk}, and Y_{3k} are components of Y_{ijk} on levels 1, 2, and 3. Here, Y_{2jk}, and Y_{3k} may be seen as random intercepts on respective levels, and Y_{1ijk} as a residual

- Some variables may not have variation over all levels. To avoid variances that are near zero which cause convergence problems specify/restrict the variation level
- WITHIN=Y, has variation on level 1, so Y_{2jk} and Y_{3k} are not in the model
- WITHIN=(level2) Y, has variation on level 1 and level 2
- WITHIN=(level3) Y, has variation on level 1 and level 3
- BETWEEN= Y, has variation on level 2 and level 3
- BETWEEN=(level2) Y, has variation on level 2
- BETWEEN=(level3) Y, has variation on level 3

Types Of Random Slopes In 3-Level Analysis

- Type 1: Defined on the level 1
  %WITHIN%
  s | y ON x;
  The random slope *s* has variance on level 2 and level 3
- Type 2: Defined on the level 2
  %BETWEEN level2%
  s | y ON x;
  The random slope *s* has variance on level 3 only
- The dependent variable can be an observed Y or a factor. The covariate X should be specified as WITHIN= for type 1 or BETWEEN=(level2) for type 2, i.e., no variation beyond the level it is used at

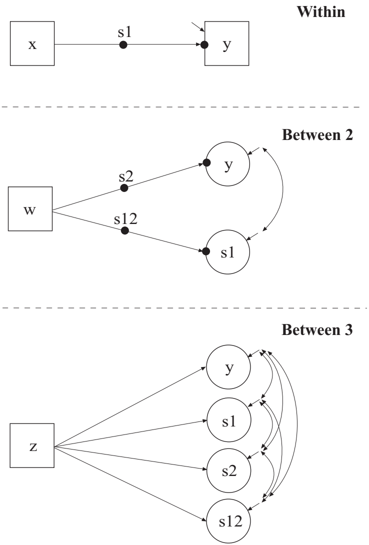
6.2 3-Level Regression

$$\begin{aligned}
Level\,1: \; y_{ijk} &= \beta_{0jk} + \beta_{1jk}\, x_{ijk} + \varepsilon_{ijk}, & (15)\\
Level\,2a: \; \beta_{0jk} &= \gamma_{00k} + \gamma_{01k}\, w_{jk} + \zeta_{0jk}, & (16)\\
Level\,2b: \; \beta_{1jk} &= \gamma_{10k} + \gamma_{11k}\, w_{jk} + \zeta_{1jk}, & (17)\\
Level\,3a: \; \gamma_{00k} &= \kappa_{000} + \kappa_{001}\, z_k + \delta_{00k}, & (18)\\
Level\,3b: \; \gamma_{01k} &= \kappa_{010} + \kappa_{011}\, z_k + \delta_{01k}, & (19)\\
Level\,3c: \; \gamma_{10k} &= \kappa_{100} + \kappa_{101}\, z_k + \delta_{10k}, & (20)\\
Level\,3d: \; \gamma_{11k} &= \kappa_{110} + \kappa_{111}\, z_k + \delta_{11k}, & (21)
\end{aligned}$$

where

- x, w, and z are covariates on the different levels
- β are level 2 random effects
- γ are level 3 random effects
- κ are fixed effects
- ε, ζ and δ are residuals on the different levels

3-Level Regression Example: UG Example 9.20



3-Level Regression Example: UG Example 9.20 Input

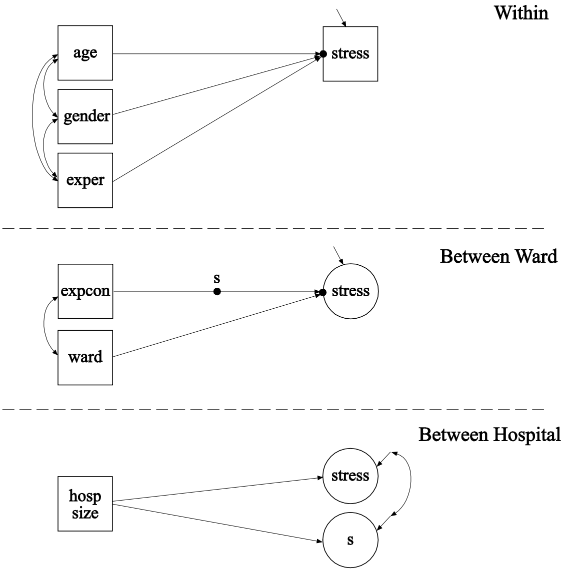
```
TITLE:      this is an example of a three-level
            regression with a continuous dependent
            variable
DATA:       FILE = ex9.20.dat;
VARIABLE:   NAMES = y x w z level2 level3;
            CLUSTER = level3 level2;
            WITHIN = x;
            BETWEEN =(level2) w (level3) z;
ANALYSIS:   TYPE = THREELEVEL RANDOM;
MODEL:
            %WITHIN%
            s1 | y ON x;
            %BETWEEN level2%
            s2 | y ON w;
            s12 | s1 ON w;
            y WITH s1;
            %BETWEEN level3%
            y ON z;
            s1 ON z;
            s2 ON z;
            s12 ON z;
            y WITH s1 s2 s12;
            s1 WITH s2 s12;
            s2 WITH s12;
OUTPUT:     TECH1 TECH8;
```

6.3 3-Level Regression: Nurses Data

Source: Hox (2010). Multilevel Analysis. Hypothetical data discussed in Section 2.4.3

- Study of stress in hospitals
- Reports from nurses working in wards nested within hospitals
- In each of 25 hospitals, 4 wards are selected and randomly assigned to experimental or control conditions (cluster-randomized trial)
- 10 nurses from each ward are given a test that measures job-related stress
- Covariates are age, experience, gender, type of ward (0=general care, 1=special care), hospital size (0=small, 1=medium, 2=large)
- Research question: Is the experimental effect different in different hospitals? - Random slope varying on level 3

3-Level Regression Example: Nurses Data



Input For Nurses Data

```
TITLE:        Nurses data from Hox (2010)
DATA:         FILE = nurses.dat;
VARIABLE:     NAMES = hospital ward wardid nurse age gender
              experience stress wardtype hospsize expcon zage
              zgender zexperience zstress zwardtyi zhospsize
              zexpcon cexpcon chospsize;
              CLUSTER = hospital wardid;
              WITHIN = age gender experience;
              BETWEEN = (hospital) hospsize (wardid) expcon wardtype;
              USEVARIABLES = stress expcon age gender experience
              wardtype hospsize;
              CENTERING = GRANDMEAN(expcon hospsize);
ANALYSIS:     TYPE = THREELEVEL RANDOM;
              ESTIMATOR = MLR;
```

Input For Nurses Data, Continued

```
MODEL:      %WITHIN%
            stress ON age gender experience;
            %BETWEEN wardid%
            s | stress ON expcon;
            stress ON wardtype;
            %BETWEEN hospital%
            s stress ON hospsize;
            s; s WITH stress;
OUTPUT:     TECH1 TECH8;
SAVEDATA:   SAVE = FSCORES;
            FILE = fs.dat;
PLOT:       TYPE = PLOT2 PLOT3;
```

Model Results For Nurses Data

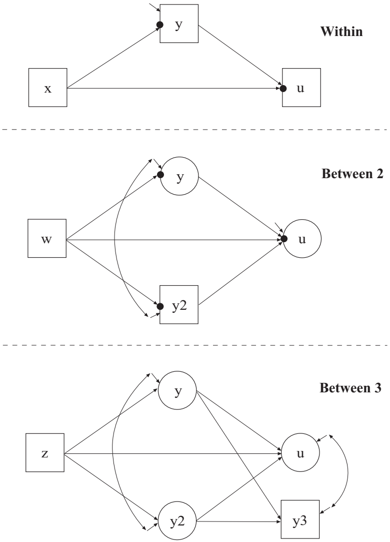
	Estimates	S.E.	Est./S.E.	Two-Tailed P-Value
WITHIN Level				
stress ON				
age	0.022	0.002	11.911	0.000
gender	-0.455	0.032	-14.413	0.000
experience	-0.062	0.004	-15.279	0.000
Residual Variances				
stress	0.217	0.011	20.096	0.000
BETWEEN wardid Level				
stress ON				
wardtype	0.053	0.076	0.695	0.487

Model Results For Nurses Data, Continued

	Estimates	S.E.	Est./S.E.	Two-Tailed P-Value
Residual Variances				
stress	0.109	0.033	3.298	0.001
BETWEEN hospital Level				
s ON				
hospsize	0.998	0.191	5.217	0.000
stress ON				
hospsize	-0.041	0.152	-0.270	0.787
s WITH				
stress	-0.036	0.058	-0.615	0.538

Model Results For Nurses Data, Continued

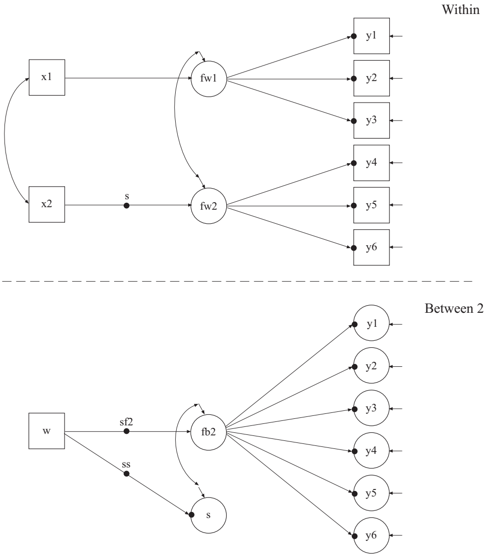
	Estimates	S.E.	Est./S.E.	Two-Tailed P-Value
Intercepts				
stress	5.753	0.102	56.171	0.000
s	-0.699	0.111	-6.295	0.000
Residual Variances				
stress	0.143	0.051	2.813	0.005
s	0.178	0.087	2.060	0.039


y
Within
x
u
y
Between 2
w
u
y2
y
Between 3
z
u
y2
y3

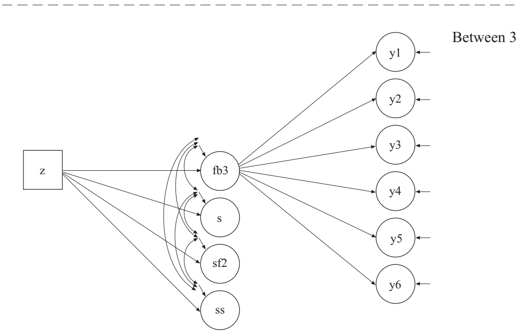
3-Level Path Analysis: UG Ex 9.21 Input

```
TITLE:     this an example of a three-level path
           analysis with a continuous and a
           categorical dependent variable
DATA:      FILE = ex9.21.dat;
VARIABLE:  NAMES = u y2 y y3 x w z level2 level3;
           CATEGORICAL  = u;
           CLUSTER = level3 level2;
           WITHIN = x;
           BETWEEN = y2 (level2) w (level3) z y3;
ANALYSIS:  TYPE = THREELEVEL;
           ESTIMATOR = BAYES;
           PROCESSORS = 2;
           BITERATIONS = (1000);
MODEL:     %WITHIN%
           u ON y x;
           y ON x;
           %BETWEEN level2%
           u ON w y y2;
           y ON w;
           y2 ON w;
           y WITH y2;
           %BETWEEN level3%
           u ON y y2;
           y ON z;
           y2 ON z;
           y3 ON y y2;
           y WITH y2;
           u WITH y3;
OUTPUT:    TECH1 TECH8;
```

6.5 3-Level MIMIC Analysis



3-Level MIMIC Analysis, Continued



3-Level MIMIC Analysis Input

```
TITLE:     this is an example of a three-level MIMIC
           model with continuous factor indicators,
           two covariates on within, one covariate on
           between level 2, one covariate on between
           level 3 with random slopes on both within
           and between level 2
DATA:      FILE = ex9.22.dat;
VARIABLE:  NAMES = y1-y6 x1 x2 w z level2 level3;
           CLUSTER = level3 level2;
           WITHIN = x1 x2;
           BETWEEN = (level2) w (level3) z;
ANALYSIS:  TYPE = THREELEVEL RANDOM;
MODEL:     %WITHIN%
           fw1 BY y1-y3;
           fw2 BY y4-y6;
           fw1 ON x1;
           s | fw2 ON x2;
           %BETWEEN level2%
           fb2 BY y1-y6;
           sf2 | fb2 ON w;
           ss | s ON w;
           fb2 WITH s;
           %BETWEEN level3%
           fb3 BY y1-y6;
           fb3 ON z;
           s ON z;
           sf2 ON z;
           ss ON z;
           fb3 WITH s sf2 ss;
           s WITH sf2 ss;
           sf2 WITH ss;
OUTPUT:    TECH1 TECH8;
```

3-Level MIMIC Analysis, Monte Carlo Input: 5 Students (14 Parameters) In 30 Classrooms (13 Parameters) In 50 Schools (28 Parameters)

```
MONTECARLO:
        NAMES = y1-y6 x1 x2 w z;
        NOBSERVATIONS = 7500;
        NREPS = 500;
        CSIZES = 50[30(5)];
        NCSIZE = 1[1];
        !SAVE = ex9.22.dat;
        WITHIN = x1 x2;
        BETWEEN = (level2) w (level3) z;
ANALYSIS:
        TYPE = THREELEVEL RANDOM;
        ESTIMATOR = MLR;
```

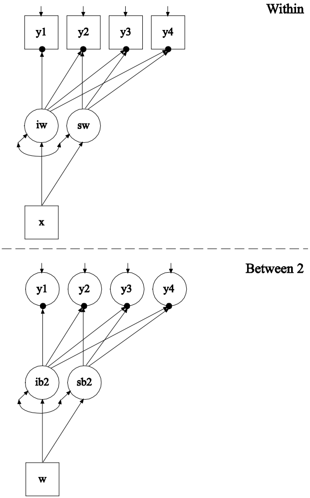
```
REPLICATION 499:
    THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES MAY NOT BE
    TRUSTWORTHY FOR SOME PARAMETERS DUE TO A NON-POSITIVE DEFINITE
    FIRST-ORDER DERIVATIVE PRODUCT MATRIX.  THIS MAY BE DUE TO THE STARTING
    VALUES BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION.  THE
    CONDITION NUMBER IS      -0.239D-16.  PROBLEM INVOLVING PARAMETER 51.

    THE NONIDENTIFICATION IS MOST LIKELY DUE TO HAVING MORE PARAMETERS THAN THE
    NUMBER OF LEVEL 3 CLUSTERS. REDUCE THE NUMBER OF PARAMETERS.

    REPLICATION 500:
    THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES MAY NOT BE
    TRUSTWORTHY FOR SOME PARAMETERS DUE TO A NON-POSITIVE DEFINITE
    FIRST-ORDER DERIVATIVE PRODUCT MATRIX.  THIS MAY BE DUE TO THE STARTING
    VALUES BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION.  THE
    CONDITION NUMBER IS      -0.190D-16.  PROBLEM INVOLVING PARAMETER 52.

    THE NONIDENTIFICATION IS MOST LIKELY DUE TO HAVING MORE PARAMETERS THAN THE
    NUMBER OF LEVEL 3 CLUSTERS. REDUCE THE NUMBER OF PARAMETERS.
```

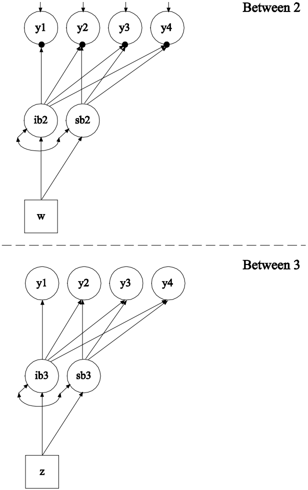
	Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Between LEVEL2 Level							
FB2 BY							
Y1	1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
Y2	1.000	0.9980	0.0236	0.0237	0.0006	0.952	1.000
Y3	1.000	0.9999	0.0237	0.0239	0.0006	0.940	1.000
Y4	1.000	0.9987	0.0271	0.0272	0.0007	0.936	1.000
Y5	1.000	1.0005	0.0265	0.0270	0.0007	0.948	1.000
Y6	1.000	0.9987	0.0277	0.0269	0.0008	0.944	1.000
FB2 WITH							
S	0.000	0.0001	0.0238	0.0222	0.0006	0.940	0.060
Residual Variances							
Y1	0.500	0.5009	0.0343	0.0338	0.0012	0.940	1.000
Y2	0.500	0.4988	0.0345	0.0338	0.0012	0.928	1.000
Y3	0.500	0.5004	0.0347	0.0336	0.0012	0.936	1.000
Y4	0.500	0.4995	0.0333	0.0339	0.0011	0.950	1.000
Y5	0.500	0.4988	0.0337	0.0337	0.0011	0.946	1.000
Y6	0.500	0.5002	0.0350	0.0339	0.0012	0.932	1.000
FB2	0.500	0.5021	0.0327	0.0321	0.0011	0.934	1.000
S	0.600	0.6018	0.0384	0.0374	0.0015	0.938	1.000

```
Between LEVEL3 Level

 FB3       BY
  Y1                1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  Y2                1.000     1.0112     0.1396     0.1372     0.0196 0.934 1.000
  Y3                1.000     1.0091     0.1608     0.1403     0.0259 0.928 1.000
  Y4                1.000     1.0063     0.1491     0.1398     0.0222 0.912 1.000
  Y5                1.000     1.0094     0.1532     0.1420     0.0235 0.920 1.000
  Y6                1.000     1.0155     0.1585     0.1418     0.0253 0.932 1.000

 FB3         ON
  Z                 0.500     0.5053     0.1055     0.0932     0.0111 0.906 1.000

 S           ON
  Z                 0.300     0.2947     0.0859     0.0791     0.0074 0.912 0.940

 SF2         ON
  Z                 0.200     0.1988     0.0834     0.0794     0.0069 0.922 0.704

 SS          ON
  Z                 0.300     0.3016     0.0863     0.0790     0.0074 0.918 0.938

 FB3       WITH
  S                 0.000     0.0018     0.0501     0.0466     0.0025 0.940 0.060
  SF2               0.000     0.0050     0.0499     0.0462     0.0025 0.944 0.056
  SS                0.000     0.0008     0.0487     0.0466     0.0024 0.932 0.068

 S         WITH
  SF2               0.000     0.0033     0.0465     0.0442     0.0022 0.938 0.062
  SS                0.000    -0.0025     0.0448     0.0438     0.0020 0.944 0.056

 SF2       WITH
  SS                0.000    -0.0008     0.0471     0.0440     0.0022 0.940 0.060
```

3-Level MIMIC Analysis, Monte Carlo Output, Continued

Intercepts							
Y1	0.500	0.4945	0.0995	0.1031	0.0099	0.966	0.996
Y2	0.500	0.4924	0.1035	0.1031	0.0108	0.932	0.992
Y3	0.500	0.4920	0.1051	0.1029	0.0111	0.942	0.998
Y4	0.500	0.4967	0.1059	0.1034	0.0112	0.940	0.998
Y5	0.500	0.4974	0.0996	0.1029	0.0099	0.946	1.000
Y6	0.500	0.4975	0.1011	0.1033	0.0102	0.950	0.996
S	0.200	0.1977	0.0837	0.0809	0.0070	0.926	0.664
SF2	1.000	1.0051	0.0867	0.0814	0.0075	0.934	1.000
SS	0.500	0.5042	0.0853	0.0808	0.0073	0.944	1.000
Residual Variances							
Y1	0.200	0.1906	0.0556	0.0506	0.0032	0.872	0.996
Y2	0.200	0.1893	0.0554	0.0499	0.0032	0.884	0.996
Y3	0.200	0.1922	0.0545	0.0504	0.0030	0.892	0.994
Y4	0.200	0.1928	0.0597	0.0502	0.0036	0.868	0.996
Y5	0.200	0.1911	0.0550	0.0507	0.0031	0.872	0.998
Y6	0.200	0.1907	0.0517	0.0504	0.0028	0.906	1.000
FB3	0.300	0.2899	0.0901	0.0842	0.0082	0.892	0.992
S	0.300	0.2885	0.0639	0.0622	0.0042	0.906	1.000
SF2	0.300	0.2905	0.0656	0.0619	0.0044	0.888	1.000
SS	0.300	0.2850	0.0673	0.0622	0.0047	0.870	1.000


Within
y1
y2
y3
y4
iw
sw
x
Between 2
y1
y2
y3
y4
ib2
sb2
w


Between 2
y1
y2
y3
y4
ib2
sb2
w
Between 3
y1
y2
y3
y4
ib3
sb3
z

3-Level Growth Analysis Input

```
TITLE:     this is an example of a three-level growth
           model with a continuous outcome and one
           covariate on each of the three levels
DATA:      FILE = ex9.23.dat;
VARIABLE:  NAMES = y1-y4 x w z level2 level3;
           CLUSTER = level3 level2;
           WITHIN = x;
           BETWEEN = (level2) w (level3) z;
ANALYSIS:  TYPE = THREELEVEL;
MODEL:     %WITHIN%
           iw sw | y1@0 y2@1 y3@2 y4@3;
           iw sw ON x;
           %BETWEEN level2%
           ib2 sb2 | y1@0 y2@1 y3@2 y4@3;
           ib2 sb2 ON w;
           %BETWEEN level3%
           ib3 sb3 | y1@0 y2@1 y3@2 y4@3;
           ib3 sb3 ON z;
           y1-y4@0;
OUTPUT:    TECH1 TECH8;
```

6.7 TYPE=THREELEVEL COMPLEX

- Asparouhov, T. and Muthén, B. (2005). Multivariate Statistical Modeling with Survey Data. Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.
- Available with ESTIMATOR=MLR when all dependent variables are continuous.
- Cluster sampling: CLUSTER=cluster4 cluster3 cluster2; For example, cluster=district school classroom;
- cluster4 nested above cluster3 nested above cluster2
- cluster4 provides information about cluster sampling of level 3 units, cluster3 is modeled as level 3, cluster2 is modeled as level 2
- cluster4 affects only the standard errors and not the point estimates, adjusts the standard error upwards for non-independence of level 3 units

TYPE=THREELEVEL COMPLEX, Continued

- Other sampling features: Stratification (nested above cluster4, 5 levels total), finite population sampling and weights
- Three weight variables for unequal probability of selection
- weight=w1; bweight=w2; b2weight=w3;

$$w3 = 1/P(\text{level 3 unit is selected})$$

$$w2 = 1/P(\text{level 2 unit is selected}|\text{the level 3 unit is selected})$$

$$w1 = 1/P(\text{level 1 unit is selected}|\text{the level 2 unit is selected})$$

- Weights are scaled to sample size at the corresponding level
- Other scaling methods possible: https://www.statmodel.com/download/Scaling3.pdf

New Multiple Imputation Methods

- Multiple imputations for three-level and cross-classified data
- Continuous and categorical variables
- H0 imputations. Estimate a three-level or cross-classified model with the Bayes estimator. Not available as H1 imputation where the imputation model is setup as unrestricted model.
- The imputation model can be an unrestricted model or a restricted model. Restricted models will be easier to estimate especially when the number of clustering units is not large
- In the input file simply add the DATA IMPUTATION command

Example Of Multiple Imputation For Three-Level Data

```
variable:
            names are y1-y10 c1 c2;
            cluster=c2 c1;
            missing=all(999);

data:       file=3imp.dat;

analysis:  type = threelevel; estimator=bayes;

data imputation:
            ndatasets = 10;
            save = 3levImp*.dat;
            impute = y1-y10;

model:
            %within%
            y1-y10*1;
            e1 by y1-y10*1; e1@1;

            %between c1%
            y1-y10*.5;
            e2 by y1-y10*1; e2@1;

            %between c2%
            y1-y10*.3;
            e3 by y1-y10*1; e3@1;
```

7. Cross-Classified Analysis: Introductory

- Regression analysis
- Path analysis (both subject and context are random modes)
- SEM
- Random items (both subject and item are random modes)
- Longitudinal analysis (both subject and time are random modes)

Cross-Classified Data Structure

Students are cross-classified by school and neighbourhood at level 2. An example with 33 students:

	School 1	School 2	School 3	School 4
Neighbourhood 1	XXXX	XX	X	X
Neighbourhood 2	X	XXXXX	XXX	XX
Neighbourhood 3	XX	XX	XXXX	XXXXXX

Source: Fielding & Goldstein (2006). Cross-classified and multiple membership structures in multilevel models: An introduction and review. Research Report RR 791, University of Birmingham.

Cross-Classified Data

- Y_{pijk} is the $p-$th observation for person i belonging to level 2 cluster j and level 3 cluster k.
- Level 2 clusters are not nested within level 3 clusters
- Examples:
  - Natural Nesting: Students performance scores are nested within students and teachers. Students are nested within schools and neighborhoods.
  - Design Nesting: Studies where observations are nested within persons and treatments/situations.
  - Complex Sampling: Observations are nested within sampling units and another variable unrelated to the sampling.
  - Generalizability theory: Items are considered a random sample from a population of items.

Why do we need to model both sets of clustering?

- Discover the true predictor/explanatory effect stemming from the clusters
- Ignoring clustering leads to incorrect standard errors
- Modeling with fixed effects leads to too many parameters and less accurate model

7.1 Cross-Classified Regression

Consider an outcome y_{ijk} for individual i nested within the cross-classification of level 2a with index j and level 2b with index k. For example, level 2a is the school an individual goes to and level 2b is the neighborhood the individual lives in. This is not a three-level structure because a school an individual goes to need not be in the neighborhood the individual lives in. Following is a simple model,

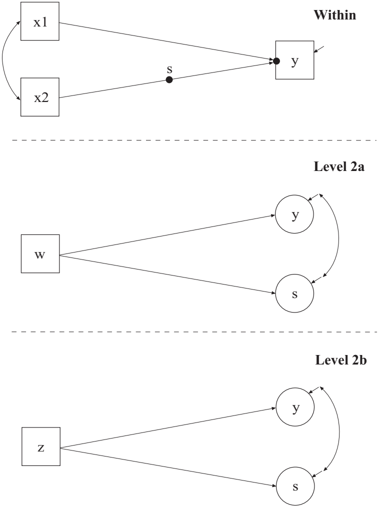
$$y_{ijk} = \beta_0 + \beta_1\, x_{ijk} + \beta_{2a\,j} + \beta_{2b\,k} + \varepsilon_{ijk}, \tag{22}$$

$$\beta_{2a\,j} = \gamma_{2a}\, w_{2a\,j} + \zeta_{2a\,j}, \tag{23}$$

$$\beta_{2b\,k} = \gamma_{2b}\, z_{2b\,k} + \zeta_{2b\,k}, \tag{24}$$

where

- x, w_{2a}, and z_{2b} are covariates on the different levels
- β_0, β_1, γ_{2a} and γ_{2b} are fixed effect coefficients on the different levels
- ε, $\beta_{2a\,j}$ and $\beta_{2b\,k}$ are random effects on the different levels


Within
x1
x2
s
y
Level 2a
w
y
s
Level 2b
z
y
s

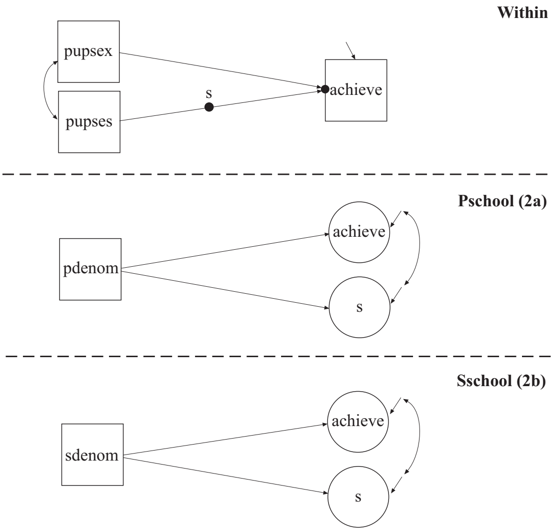
```
TITLE:     this is an example of a two-level
           regression for a continuous dependent
           variable using cross-classified data
DATA:      FILE = ex9.24.dat;
VARIABLE:  NAMES = y x1 x2 w z level2a level2b;
           CLUSTER = level2b level2a;
           WITHIN = x1 x2;
           BETWEEN = (level2a) w (level2b) z;
ANALYSIS:  TYPE = CROSSCLASSIFIED RANDOM;
           ESTIMATOR = BAYES;
           PROCESSORS = 2;
           BITERATIONS = (2000);
MODEL:     %WITHIN%
           y ON x1;
           s | y ON x2;
           %BETWEEN level2a%
           y ON w;
           s ON w;
           y WITH s;
           %BETWEEN level2b%
           y ON z;
           s ON Z;
           y WITH s;
OUTPUT:    TECH1 TECH8;
```

7.3 Cross-Classified Regression: Pupcross Data

Hox (2010). Multilevel Analysis. Second edition. Chapter 9.1

- 1000 pupils, attending 100 different primary schools, going on to 30 secondary schools
- Outcome: Achievement measured in secondary school
- x covariate: pupil gender (0=male, 1=female), pupil ses
- w_{2a} covariate: pdenom (0=public, 1=denom); primary school denomination
- z_{2b} covariate: sdenom (0=public, 1=denom); secondary school denomination

Cross-Classified Modeling Of Pupcross Data



```
TITLE:      Pupcross: No covariates
DATA:       FILE = pupcross.dat;
VARIABLE:   NAMES = pupil pschool sschool achieve pupsex pupses
            pdenom sdenom;
            USEVARIABLES = achieve;
            CLUSTER = pschool sschool;
ANALYSIS:   ESTIMATOR = BAYES;
            TYPE = CROSSCLASSIFIED;
            PROCESSORS = 2;
            FBITER = 5000;
MODEL:      %WITHIN%
            achieve;
            %BETWEEN pschool%
            achieve;
            %BETWEEN sschool%
            achieve;
OUTPUT:     TECH1 TECH8;
```

Cluster information for SSCHOOL		
Size (s)	Cluster ID with Size s	
20	9	
21	20	
22	12	
23	24	
24	15	
26	3	17
27	1	30
28	23	
30	5	
31	26	25
32	2	
33	8	13
34	4	18
35	29	
37	27	11
39	22	19
41	16	
42	21	7
45	14	
46	10	
47	28	

Cluster information for PSCHOOL								
Size (s)	Cluster ID with Size s							
10	50							
12	43							
13	41	24						
15	47	23	5	22				
16	30	9						
17	7	26	38					
18	1	3	6	45	14	28		
19	29	17	49	35	21	20		
20	16	2						
21	40	32	46	11	19	13	4	39
22	34	27						
23	15	18						
24	25	44	37					
25	36	31	10					
26	8							
27	42							
29	48	12						
31	33							

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.%	Upper 2.5%	Significance
WITHIN level						
Variances						
achieve	0.513	0.024	0.000	0.470	0.564	*
BETWEEN sschool level						
Variances						
achieve	0.075	0.028	0.000	0.040	0.147	*
BETWEEN pschool level						
Means						
achieve	6.341	0.084	0.000	6.180	6.510	*
Variances						
achieve	0.183	0.046	0.000	0.116	0.294	*

```
TITLE:      Pupcross: Adding pupil gender and ses
DATA:       FILE = pupcross.dat;
VARIABLE:   NAMES = pupil pschool sschool achieve pupsex pupses
            pdenom sdenom;
            USEVARIABLES = achieve pupsex pupses;
            CLUSTER = pschool sschool;
            WITHIN = pupsex pupses;
ANALYSIS:   ESTIMATOR = BAYES;
            TYPE = CROSSCLASSIFIED;
            PROCESSORS = 2;
            FBITER = 5000;
MODEL:      %WITHIN%
            achieve ON pupsex pupses;
            %BETWEEN pschool%
            achieve;
            %BETWEEN school%
            achieve;
OUTPUT:     TECH1 TECH8;
```

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.%	95% C.I. Upper 2.5%	Significance
WITHIN level						
achieve ON						
pupsex	0.262	0.046	0.000	0.171	0.353	*
pupses	0.114	0.016	0.000	0.081	0.145	*
Residual variances						
achieve	0.477	0.022	0.000	0.434	0.523	*
BETWEEN sschool level						
Variances						
achieve	0.073	0.028	0.000	0.038	0.145	*
BETWEEN pschool level						
Means						
achieve	5.757	0.109	0.000	5.539	5.975	*
Variances						
achieve	0.183	0.046	0.000	0.116	0.297	*

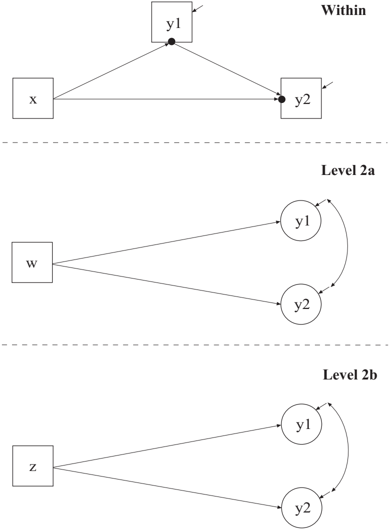
```
TITLE:       Pupil gender and ses with random ses slope for primary schools
VARIABLE:    NAMES = pupil pschool sschool achieve pupsex pupses
             pdenom sdenom;
             USEVARIABLES = achieve pupsex pupses;
             CLUSTER = pschool sschool;
             WITHIN = pupsex pupses;
ANALYSIS:    ESTIMATOR = BAYES;
             TYPE = CROSSCLASSIFIED RANDOM ;
             PROCESSORS = 2; FBITER = 5000;
MODEL:       %WITHIN%
             achieve ON pupsex;
             s | achieve ON pupses;
             %BETWEEN PSCHOOL%
             achieve;
             s;
             %BETWEEN SSCHOOL%
             achieve;
             s@0;
OUTPUT:      TECH1 TECH8;
```

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.%	Upper 2.5%	Significance
WITHIN level						
achieve ON						
pupsex	0.253	0.045	0.000	0.163	0.339	*
Residual variances						
achieve	0.465	0.022	0.000	0.424	0.510	*
BETWEEN sschool level						
Variances						
achieve	0.071	0.027	0.000	0.038	0.140	*
s	0.000	0.000	0.000	0.000	0.000	
BETWEEN pschool level						
Means						
achieve	5.758	0.105	0.000	5.557	5.964	*
s	0.116	0.019	0.000	0.077	0.153	*
Variances						
achieve	0.110	0.045	0.000	0.042	0.216	*
s	0.006	0.002	0.000	0.002	0.011	*

```
TITLE:      Pupil gender and ses plus pschool pdenom
VARIABLE:   NAMES = pupil pschool sschool achieve pupsex pupses
            pdenom sdenom;
            USEVARIABLES = achieve pupsex pupses pdemon; !sdenom;
            CLUSTER = pschool sschool;
            WITHIN = pupsex pupses;
            BETWEEN = (pschool) pdenom; ! (sschool) sdenom;
ANALYSIS:   ESTIMATOR = BAYES;
            TYPE = CROSSCLASSIFIED;
            PROCESSORS = 2; FBITER = 5000;
MODEL:      %WITHIN%
            achieve ON pupsex pupses;
            %BETWEEN PSCHOOL%
            achieve ON pdenom;
            %BETWEEN SSCHOOL%
            achieve; ! ON sdenom;
OUTPUT:     TECH1 TECH8;
PLOT:       TYPE = PLOT3;
```

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.%	Upper 2.5%	Significance
WITHIN level						
achieve ON						
pupsex	0.261	0.047	0.000	0.168	0.351	*
pupses	0.113	0.016	0.000	0.080	0.143	*
Residual variances						
achieve	0.477	0.023	0.000	0.436	0.522	*
BETWEEN sschool level						
Variances						
achieve	0.073	0.028	0.000	0.038	0.145	*
BETWEEN pschool level						
achieve ON						
pdenom	0.207	0.131	0.058	-0.053	0.465	
Intercepts						
achieve	5.643	0.136	0.000	5.375	5.912	*
Residual variances						
achieve	0.175	0.045	0.000	0.112	0.288	*

7.4 Cross-Classified Path Analysis: UG Example 9.25



```
TITLE:      this is an example of a two-level path
            analysis with continuous dependent
            variables using cross-classified data
DATA:       FILE =   ex9.25.dat;
VARIABLE:   NAMES = y1 y2 x w z level2a level2b;
            CLUSTER = level2b level2a;
            WITHIN = x;
            BETWEEN = (level2a) w (level2b) z;
ANALYSIS:   TYPE = CROSSCLASSIFIED;
             ESTIMATOR = BAYES;
            PROCESSORS = 2;
MODEL:      %WITHIN%
            y2 ON y1 x;
            y1 ON x;
            %BETWEEN level2a%
            y1-y2 ON w;
            y1 WITH y2;
            %BETWEEN level2b%
            y1-y2 ON z;
            y1 WITH y2;
OUTPUT:     TECH1 TECH8;
```

Advanced topics:

- 2-mode path analysis
- Cross-classified SEM
- Random item IRT

Gonzalez, de Boeck, Tuerlinckx (2008). A double-structure structural equation model for three-mode data. Psychological Methods, 13, 337-353.

- A population of situations that might elicit negative emotional responses
- 11 situations (e.g. blamed for someone else's failure after a sports match, a fellow student fails to return your notes the day before an exam, you hear that a friend is spreading gossip about you) viewed as randomly drawn from a population of situations
- 4 binary responses: Frustration, antagonistic action, irritation, anger
- n=679 high school students
- Level 2 cluster variables are situations and students
- 1 observation for each pair of clustering units

2-Mode Path Analysis: Random Contexts In Gonzalez Et Al.

Research questions: Which of the relationships below are significant? Are the relationships the same on the situation level as on the subject level?

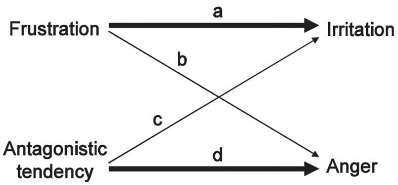
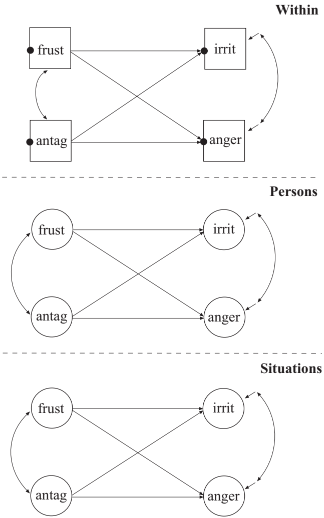


Figure 3. Graphical representation of the research questions. *a*, *b*, *c*, and *d* are effect parameters.

2-Mode Path Analysis: Random Contexts In Gonzalez Et Al.



2-Mode Path Analysis Input

```
VARIABLE:   NAMES = frust antag irrit anger student situation;
            CLUSTER = situation student;
            CATEGORICAL = frust antag irrit anger;
DATA:       FILE = gonzalez.dat;
ANALYSIS:   TYPE = CROSSCLASSIFIED;
            ESTIMATOR = BAYES;
            BITERATIONS = (10000);
MODEL:      %WITHIN%
            irrit anger ON frust antag;
            irrit WITH anger;
            frust WITH antag;
            %BETWEEN student%
            irrit ON frust (1);
            anger ON frust (2);
            irrit ON antag (3);
            anger ON antag (4);
            irrit; anger; irrit WITH anger;
            frust; antag; frust WITH antag;
```

2-Mode Path Analysis Input, Continued

```
                  %BETWEEN situation%
                  irrit ON frust (1);
                  anger ON frust (2);
                  irrit ON antag (3);
                  anger ON antag (4);
                  irrit; anger; irrit WITH anger;
                  frust; antag; frust WITH antag;
OUTPUT:           TECH8 TECH9 STDY;
PLOT:             TYPE = PLOT2;
```

8.2 2-Mode Path Analysis: Monte Carlo Simulation Using The Gonzalez Model

M is the number of cluster units for both between levels, β is the common slope, ψ is the within-level correlation, τ is the binary outcome threshold. Table gives bias (coverage).

Para	M=10	M=20	M=30	M=50	M=100
β_1	0.13(0.92)	0.05(0.89)	0.00(0.97)	0.01(0.92)	0.01(0.94)
$\psi_{2,11}$	0.11(1.00)	0.06(0.96)	0.01(0.98)	0.00(0.89)	0.02(0.95)
$\psi_{2,12}$	0.15(0.97)	0.06(0.92)	0.05(0.97)	0.03(0.87)	0.01(0.96)
τ_1	0.12(0.93)	0.01(0.93)	0.00(0.90)	0.03(0.86)	0.00(0.91)

Small biases for $M = 10$. Due to parameter equalities information is combined from both clustering levels. Adding unconstrained level 1 model: tetrachoric correlation matrix.

- General SEM model: 2-way ANOVA. Y_{pijk} is the $p-$th variable for individual i in cluster j and cross cluster k

$$Y_{pijk} = Y_{1pijk} + Y_{2pj} + Y_{3pk}$$

- 3 sets of structural equations - one on each level

$$Y_{1ijk} = \nu + \Lambda_1 \eta_{ijk} + \varepsilon_{ijk}$$

$$\eta_{ijk} = \alpha + B_1 \eta_{ijk} + \Gamma_1 x_{ijk} + \xi_{ijk}$$

$$Y_{2j} = \Lambda_2 \eta_j + \varepsilon_j$$

$$\eta_j = B_2 \eta_j + \Gamma_2 x_j + \xi_j$$

$$Y_{3k} = \Lambda_3 \eta_k + \varepsilon_k$$

$$\eta_k = B_3 \eta_k + \Gamma_3 x_k + \xi_k$$

Cross - Classified SEM

- The regression coefficients on level 1 can be a random effects from each of the two clustering levels: combines cross-classified SEM and cross classified HLM
- Bayesian MCMC estimation: used as a frequentist estimator.
- Easily extends to categorical variables.
- ML estimation possible only when one of the two level of clustering has small number of units.

- 1 factor at the individual level and 1 factor at each of the clustering levels, 5 indicator variables on the individual level

$$y_{pijk} = \mu_p + \lambda_{1,p} f_{1,ijk} + \lambda_{2,p} f_{2,j} + \lambda_{3,p} f_{3,k} + \varepsilon_{2,pj} + \varepsilon_{3,pk} + \varepsilon_{1,pijk}$$

- M level 2 clusters. M level 3 clusters. 1 unit within each cluster intersection. More than 1 unit is possible. Zero units possible: sparse tables
- Monte Carlo simulation: Estimation takes less than 1 min per replication

Cross-Classified Model Example 1: Factor Model Results

Table: Absolute bias and coverage for cross-classified factor analysis model

Param	M=10	M=20	M=30	M=50	M=100
$\lambda_{1,1}$	0.07(0.92)	0.03(0.89)	0.01(0.95)	0.00(0.97)	0.00(0.91)
$\theta_{1,1}$	0.05(0.96)	0.00(0.97)	0.00(0.95)	0.00(0.99)	0.00(0.94)
$\lambda_{2,p}$	0.21(0.97)	0.11(0.94)	0.10(0.93)	0.06(0.94)	0.00(0.92)
$\theta_{2,p}$	0.24(0.99)	0.10(0.95)	0.04(0.92)	0.05(0.94)	0.02(0.96)
$\lambda_{3,p}$	0.45(0.99)	0.10(0.97)	0.03(0.99)	0.01(0.92)	0.03(0.97)
$\theta_{3,p}$	0.75(1.00)	0.25(0.98)	0.15(0.97)	0.12(0.98)	0.05(0.92)
μ_p	0.01(0.99)	0.04(0.98)	0.01(0.97)	0.05(0.99)	0.00(0.97)

Perfect coverage. Level 1 parameters estimated very well. Biases when the number of clusters is small $M = 10$. Weakly informative priors can reduce the bias for small number of clusters.

8.5 Cross-Classified Models: Types Of Random Effects

- Type 1: Random slope.
  %WITHIN%
  s | y ON x;
  s has variance on both crossed levels. Dependent variable can be within-level factor. Covariate x should be on the WITHIN = list.
- Type 2: Random loading.
  %WITHIN%
  s | f BY y;
  s has variance on both crossed levels. *f* is a within-level factor. The dependent variable can be a within-level factor.
- Type 3: Crossed random loading.
  %BETWEEN level2a%
  s | f BY y;
  s has variance on crossed level 2b and is defined on crossed level 2a. *f* is a level 2a factor, *s* is a level 2b factor. This is a way to use the interaction term $s \cdot f$.

8.6 Random Items, Generalizability Theory

- Items are random samples from a population of items.
- The same or different items may be administered to individuals.
- Suited for computer generated items and adaptive testing.
- 2-parameter IRT model

$$P(Y_{ij} = 1) = \Phi(a_j \theta_i + b_j)$$

- $a_j \sim N(a, \sigma_a)$, $b_j \sim N(b, \sigma_b)$: random discrimination and difficulty parameters
- The ability parameter is $\theta_i \sim N(0,1)$
- Cross-classified model. Nested within items and individuals. 1 or 0 observation in each cross-classified cell.
- Interaction of two latent variables: a_j and θ_i: Type 3 crossed random loading
- The model has only 4 parameters - much more parsimonious than regular IRT models.

Random Item 2-Parameter IRT Model Setup

```
VARIABLE:
        NAMES = u item individual;
        CLUSTER = item individual;
        CATEGORICAL = u;
ANALYSIS:
        TYPE = CROSS RANDOM;
        ESTIMATOR = BAYES;
MODEL:
        %WITHIN%

        %BETWEEN individual%
        s | f BY u;
        f@1 u@0;
        %BETWEEN item%
        u s;
```

8.7 Random Item 2-Parameter IRT: TIMMS Example

- Fox (2010) Bayesian Item Response Theory. Section 4.3.3. Dutch Six Graders Math Achievement. Trends in International Mathematics and Science Study: TIMMS 2007
- 8 test items, 478 students

Table: Random 2-parameter IRT

parameter	estimate	SE
average discrimination a	0.752	0.094
average difficulty b	0.118	0.376
variation of discrimination a	0.050	0.046
variation of difficulty b	1.030	0.760

- 8 items means that there are only 8 clusters on the %between item% level and therefore the variance estimates at that level are affected by their priors. If the number of clusters is less than 10 or 20 there is prior dependence in the variance parameters.

- Using factor scores estimation we can estimate item specific parameter and SE using posterior mean and posterior standard deviation.

Table: Random 2-parameter IRT item specific parameters

item	discrimination	SE	difficulty	SE
Item 1	0.797	0.11	-1.018	0.103
Item 2	0.613	0.106	-0.468	0.074
Item 3	0.905	0.148	-1.012	0.097
Item 4	0.798	0.118	-1.312	0.106
Item 5	0.538	0.099	0.644	0.064
Item 6	0.808	0.135	0.023	0.077
Item 7	0.915	0.157	0.929	0.09
Item 8	0.689	0.105	1.381	0.108

Random Item 2-Parameter IRT: TIMMS Example, Comparison With ML

Table: Random 2-parameter IRT item specific parameters

item	Bayes random discrimination	Bayes random SE	ML fixed discrimination	ML fixed SE
Item 1	0.797	0.110	0.850	0.155
Item 2	0.613	0.106	0.579	0.102
Item 3	0.905	0.148	0.959	0.170
Item 4	0.798	0.118	0.858	0.172
Item 5	0.538	0.099	0.487	0.096
Item 6	0.808	0.135	0.749	0.119
Item 7	0.915	0.157	0.929	0.159
Item 8	0.689	0.105	0.662	0.134

- Bayes random estimates are shrunk towards the mean and have smaller standard errors: shrinkage estimate

Random Item 2-Parameter IRT: TIMMS Example, Continued

- One can add a predictor for a person's ability. For example adding gender as a predictor yields an estimate of 0.283 (0.120), saying that males have a significantly higher math mean.
- Predictors for discrimination and difficulty random effects, for example, geometry indicator.
- More parsimonious model can yield more accurate ability estimates.

8.8 Random Item Rasch IRT Example

- De Boeck (2008) Random item IRT models
- 24 verbal aggression items, 316 persons

$$P(Y_{ij} = 1) = \Phi(\theta_i + b_j)$$

$$b_j \sim N(b, \sigma)$$

$$\theta_i \sim N(0, \tau)$$

Table: Random Rasch IRT - variance decomposition

parameter	person τ	item σ	error
estimates(SE)	1.89(0.19)	1.46(0.53)	2.892
variance explained	30%	23%	46%

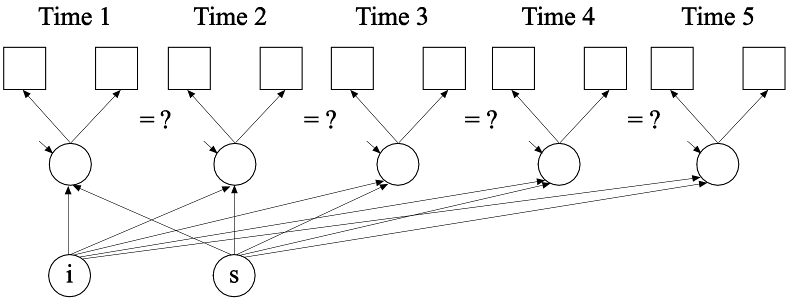
Random Item Rasch IRT Example: Simple Model Specification

```
MODEL:
        %WITHIN%

        %BETWEEN person%
        y;

        %BETWEEN item%
        y;
```

9. Advances In Longitudinal Analysis

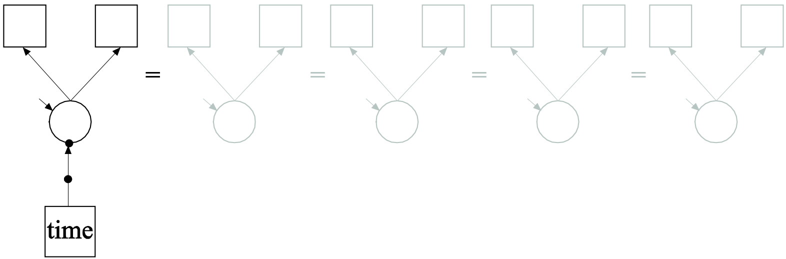
- An old dilemma
- Two new solutions

Categorical Items, Wide Format, Single-Level Approach



Single-level analysis with $p \times T = 2 \times 5 = 10$ variables, $T = 5$ factors.

- ML hard and impossible as T increases (numerical integration)
- WLSMV possible but hard when $p \times T$ increases and biased unless attrition is MCAR or multiple imputation is done first
- Bayes possible
- Searching for partial measurement invariance is cumbersome



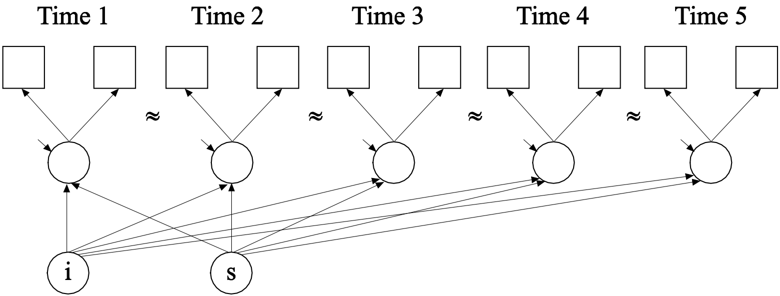
Two-level analysis with $p = 2$ variables, 1 within-factor, 2-between factors, **assuming full measurement invariance across time**.

- ML feasible
- WLSMV feasible (2-level WLSMV)
- Bayes feasible

Measurement Invariance Across Time

- Both old approaches have problems
  - Wide, single-level approach easily gets significant non-invariance and needs many modifications
  - Long, two-level approach has to assume invariance
- New solution no. 1, suitable for small to medium number of time points
  - A new wide, single-level approach where time is a fixed mode
- New solution no. 2, suitable for medium to large number of time points
  - A new long, two-level approach where time is a random mode
  - No limit on the number of time points

New Solution No. 1: Wide Format, Single-Level Approach



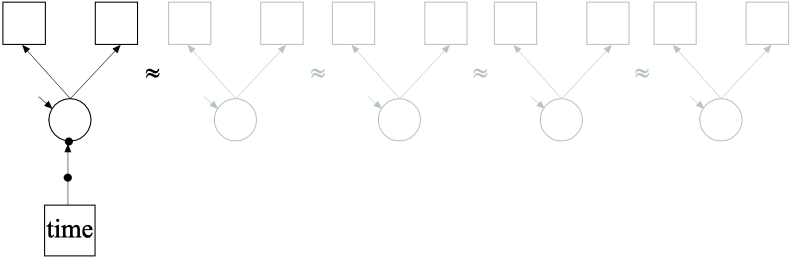
Single-level analysis with $p \times T = 2 \times 5 = 10$ variables, $T = 5$ factors.

- Bayes ("BSEM") using approximate measurement invariance, still identifying factor mean and variance differences across time

Measurement Invariance Across Time

- New solution no. 2, time is a random mode
- A new long, two-level approach
  - Best of both worlds: Keeping the limited number of variables of the two-level approach without having to assume invariance

New Solution No. 2: Long Format, Two-Level Approach



Two-level analysis with $p = 2$ variables.

- Bayes twolevel random approach with random measurement parameters and random factor means and variances using Type=Crossclassified: Clusters are time and person

9.1 Aggressive-Disruptive Behavior In The Classroom

Randomized field experiment in Baltimore public schools with a classroom-based intervention aimed at reducing aggressive-disruptive behavior among elementary school students (Ialongo et al., 1999).

This analysis:

- Cohort 1
- 9 binary items at 8 time points, Grade 1 - Grade 7
- $n = 1174$

Aggressive-Disruptive Behavior In The Classroom: ML Versus BSEM

- Traditional ML analysis
  - 8 dimensions of integration
  - Computing time: 25:44 with Integration = Montecarlo(5000)
  - Increasing the number of time points makes ML impossible
- BSEM analysis with approximate measurement invariance across time
  - 156 parameters
  - Computing time: 4:01
  - Increasing the number of time points has relatively less impact

BSEM Input Excerpts For Aggressive-Disruptive Behavior

```
            USEVARIABLES = stub1f-tease7s;
            CATEGORICAL = stub1f-tease7s;
            MISSING = ALL (999);
DEFINE:     CUT stub1f-tease7s (1.5);
ANALYSIS:   ESTIMATOR = BAYES;
            PROCESSORS = 2;
MODEL:      f1f by stub1f-tease1f* (lam11-lam19);
            f1s by stub1s-tease1s* (lam21-lam29);
            f2s by stub2s-tease2s* (lam31-lam39);
            f3s by stub3s-tease3s* (lam41-lam49);
            f4s by stub4s-tease4s* (lam51-lam59);
            f5s by stub5s-tease5s* (lam61-lam69);
            f6s by stub6s-tease6s* (lam71-lam79);
            f7s by stub7s-tease7s* (lam81-lam89);
            f1f@1;
```

```
              [stub1f$1-tease1f$1] (tau11-tau19);
              [stub1s$1-tease1s$1] (tau21-tau29);
              [stub2s$1-tease2s$1] (tau31-tau39);
              [stub3s$1-tease3s$1] (tau41-tau49);
              [stub4s$1-tease4s$1] (tau51-tau59);
              [stub5s$1-tease5s$1] (tau61-tau69);
              [stub6s$1-tease6s$1] (tau71-tau79);
              [stub7s$1-tease7s$1] (tau81-tau89);
              [f1f-f7s@0];
              i s q | f1f@0 f1s@0.5 f2s@1.5 f3s@2.5 f4s@3.5
              f5s@4.5 f6s@5.5 f7s@6.5;
              q@0;
MODEL
PRIORS:       DO(1,9) DIFF(lam1#-lam8#) ~ N(0,.01);
              DO(1,9) DIFF(tau1#-tau8#) ~ N(0,.01);
OUTPUT:       TECH1 TECH8;
```

Estimates For Aggressive-Disruptive Behavior

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	95% C.I. Upper 2.5%	
Means						
I	0.000	0.000	1.000	0.000	0.000	
S	0.238	0.068	0.000	0.108	0.366	*
Q	-0.022	0.011	0.023	-0.043	0.000	*
Variances						
I	9.258	2.076	0.000	6.766	14.259	*
S	0.258	0.068	0.000	0.169	0.411	*
Q	0.001	0.000	0.000	0.001	0.001	

Estimates For Aggressive-Disruptive Behavior, Continued

		Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%	
F1F BY							
	STUB1F	0.428	0.048	0.000	0.338	0.522	*
	BKRULE1F	0.587	0.068	0.000	0.463	0.716	*
	HARMO1F	0.832	0.082	0.000	0.677	0.985	*
	BKTHIN1F	0.671	0.067	0.000	0.546	0.795	*
	YELL1F	0.508	0.055	0.000	0.405	0.609	*
	TAKEP1F	0.717	0.072	0.000	0.570	0.839	*
	FIGHT1F	0.480	0.052	0.000	0.385	0.579	*
	LIES1F	0.488	0.054	0.000	0.386	0.589	*
	TEASE1F	0.503	0.055	0.000	0.404	0.608	*
...							
F7S BY							
	STUB7S	0.360	0.049	0.000	0.273	0.458	*
	BKRULE7S	0.512	0.068	0.000	0.392	0.654	*
	HARMO7S	0.555	0.074	0.000	0.425	0.716	*
	BKTHIN7S	0.459	0.063	0.000	0.344	0.581	*
	YELL7S	0.525	0.062	0.000	0.409	0.643	*
	TAKEP7S	0.500	0.069	0.000	0.372	0.634	*
	FIGHT7S	0.515	0.067	0.000	0.404	0.652	*
	LIES7S	0.520	0.070	0.000	0.392	0.653	*
	TEASE7S	0.495	0.064	0.000	0.378	0.626	*

Displaying Non-Invariant Items: Time Points With Significant Differences Compared To The Mean (V = 0.01)

Item	Loading	Threshold
stub	3	1, 2, 3, 6, 8
bkrule	-	5, 8
harmo	1, 8	2, 8
bkthin	1, 2, 3, 7, 8	2, 8
yell	2, 3, 6	-
takep	1, 2, 5	1, 2, 5
fight	1, 5	1, 4
lies	-	-
tease	-	1, 4, 8

9.2 Cross-Classified Analysis Of Longitudinal Data

- Observations nested within time and subject
- A large number of time points can be handled via Bayesian analysis
- A relatively small number of subjects is needed

Intensive Longitudinal Data

- Time intensive data: More longitudinal data are collected where very frequent observations are made using new tools for data collection. Walls & Schafer (2006)
- Typically multivariate models are developed but if the number of time points is large these models will fail due to too many variables and parameters involved
- Factor analysis models will be unstable over time. Is it lack of measurement invariance or insufficient model?
- Random loading and intercept models can take care of measurement and intercept invariance. A problem becomes an advantage.
- Random loading and intercept models produce more accurate estimates for the loadings and factors by borrowing information over time
- Random loading and intercept models produce more parsimonious model

Generating The Data For Ex9.27

```
TITLE:        this is an example of longitudinal modeling using a
              cross-classified data approach where observations are
              nested within the cross-classification of time and subjects
MONTECARLO:
              NAMES = y1-y3;
              NOBSERVATIONS = 7500;
              NREPS = 1;
              CSIZES = 75[100(1)];! 75 subjects, 100 time points
              NCSIZE = 1[1];
              WITHIN = (level2a) y1-y3;
              SAVE = ex9.27.dat;
ANALYSIS:
              TYPE = CROSS RANDOM;
              ESTIMATOR = BAYES;
              PROCESSORS = 2;
```

```
MODEL
POPULATION:
        %WITHIN%
        s1-s3 | f by y1-y3;
        f@1;
        y1-y3*1.2; [y1-y3@0];
        %BETWEEN level2a% ! across time variation
        s1-s3*0.1;
        [s1-s3*1.3];
        y1-y3*.5;
        [y1-y3@0];
        %BETWEEN level2b% ! across subjects variation
        f*1; [f*.5];
        s1-s3@0; [s1-s3@0];
```

```
TITLE:     this is an example of a multiple indicator
           growth model with random intercepts and
           factor loadings using cross-classified
           data
DATA:      FILE = ex9.27.dat;
VARIABLE:  NAMES = y1-y3 time subject;
           USEVARIABLES = y1-y3 timescor;
           CLUSTER = subject time;
           WITHIN = timescor (time) y1-y3;
DEFINE:    timescor = (time-1)/100;
ANALYSIS:  TYPE = CROSSCLASSIFIED RANDOM;
           ESTIMATOR = BAYES;
           PROCESSORS = 2;
           BITERATIONS = (1000);
MODEL:     %WITHIN%
           s1-s3 | f BY y1-y3;
           f@1;
           s | f ON timescor; !slope growth factor s
           y1-y3; [y1-y3@0];
           %BETWEEN time% ! time variation
           s1-s3; [s1-s3]; ! random loadings
           y1-y3; [y1-y3@0]; ! random intercepts
           s@0; [s@0];
           %BETWEEN subject% ! subject variation
           f; [f]; ! intercept growth factor f
           s1-s3@0; [s1-s3@0];
           s; [s]; ! slope growth factor s
OUTPUT:    TECH1 TECH8;
```

9.5 Cross-Classified Analysis Of Aggressive-Disruptive Behavior In The Classroom

- Teacher-rated measurement instrument capturing aggressive-disruptive behavior among a sample of U.S. students in Baltimore public schools (Ialongo et al., 1999).
- The instrument consists of 9 items scored as 0 (almost never) through 6 (almost always)
- A total of 1174 students are observed in 41 classrooms from Fall of Grade 1 through Grade 6 for a total of 8 time points
- The multilevel (classroom) nature of the data is ignored in the current analyses
- The item distribution is very skewed with a high percentage in the Almost Never category. The items are therefore dichotomized into Almost Never versus the other categories combined
- We analyze the data on the original scale as continuous variables and also the dichotomized scale as categorical

Aggressive-Disruptive Behavior Example Continued

- For each student a 1-factor analysis model is estimated with the 9 items at each time point
- Let Y_{pit} be the $p-$th item for individual i at time t
- We use cross-classified SEM. Observations are nested within individual and time.
- Although this example uses only 8 time points the models can be used with any number of time points.

- Model 1: Two-level factor model with intercept non-invariance across time

$$Y_{pit} = \mu_p + \zeta_{pt} + \xi_{pi} + \lambda_p \eta_{it} + \varepsilon_{pit}$$

- μ_p, λ_p are model parameters, $\varepsilon_{pit} \sim N(0, \theta_{w,p})$ is the residual
- $\zeta_{pt} \sim N(0, \sigma_p)$ is a random effect to accommodate intercept non-invariance across time
- To correlate the factors η_{it} within individual i

$$\eta_{it} = \eta_{b,i} + \eta_{w,it}$$

- $\eta_{b,i} \sim N(0, \psi)$ and $\eta_{w,it} \sim N(0, 1)$. The variance is fixed to 1 to identify the scale in the model
- $\xi_{pi} \sim N(0, \theta_{b,p})$ is a between level residual in the between level factor model
- Without the random effect ζ_{pt} this is just a standard two-level factor model

Aggressive-Disruptive Behavior Example Continued: Model 1 Setup

MODEL:

```
%WITHIN%
f BY y1-y9*1 (11-19);
f@1;

%BETWEEN t1%
y1-y9;

%BETWEEN id%
y1-y9;
fb BY y1-y9*1 (11-19);
```

Aggressive-Disruptive Behavior Example Cont'd: Model 2

- Model 2: Adding latent growth model for the factor

$$\eta_{it} = \alpha_i + \beta_i \cdot t + \eta_{w,it}$$

- $\alpha_i \sim N(0, v_\alpha)$ is the intercept and $\beta_i \sim N(\beta, v_\beta)$ is the slope. For identification purposes again $\eta_{w,it} \sim N(0,1)$
- The model looks for developmental trajectory across time for the aggressive-disruptive behavior factor

Aggressive-Disruptive Behavior Example Continued: Model 2 Setup

MODEL:

```
! s = beta, fb = alpha
%WITHIN%
f BY y1-y9*1 (11-19);
f@1;
s | f ON time;

%BETWEEN t1%
y1-y9;
s@0; [s@0];

%BETWEEN id%
y1-y9;
fb BY y1-y9*1 (11-19);
s*1; [s*0];
```

Aggressive-Disruptive Behavior Example Cont'd: Model 3

- Model 3: Adding measurement non-invariance
- Replace the fixed loadings λ_p with random loadings $\lambda_{pt} \sim N(\lambda_p, w_p)$
- The random loadings accommodate measurement non-invariance across time
- All models can be estimated for continuous and categorical scale data

Aggressive-Disruptive Behavior Example Continued: Model 3 Setup

MODEL:

```
%WITHIN%
s1-s9 | f BY y1-y9;
f@1;
s | f ON time;
%BETWEEN t1%
y1-y9;
f@0; [f@0];
s@0; [s@0];
s1-s9*1; [s1-s9*1];
%BETWEEN id%
y1-y9;
f*1; [f@0];
s*1; [s*0];
s1-s9@0; [s1-s9@0];
```

Aggressive-Disruptive Behavior Example Continued: Model 3 Results For Continuous Analysis

```
Within Level
 Residual Variances
    Y1                1.073      0.022      0.000      1.029      1.119
    Y9                0.630      0.014      0.000      0.604      0.658
    F                 1.000      0.000      0.000      1.000      1.000

Between ID Level
 Variances
    Y1                0.146      0.016      0.000      0.118      0.180
    Y9                0.052      0.009      0.000      0.035      0.068
    F                 1.316      0.080      0.000      1.172      1.486
    S                 0.026      0.003      0.000      0.020      0.032

Between T1 Level
 Means
    Y1                1.632      0.120      0.000      1.377      1.885
    Y9                1.232      0.096      0.000      1.044      1.420
    S1                0.679      0.023      0.000      0.640      0.732
    S9                0.705      0.043      0.000      0.628      0.797
 Variances
    Y1                0.080      0.138      0.000      0.025      0.372
    Y9                0.047      0.109      0.000      0.017      0.266
    S1                0.002      0.004      0.000      0.000      0.013
    S9                0.010      0.079      0.000      0.003      0.052
```

- Model 4: Adding measurement non-invariance also across individuals
- Replace the loadings λ_{pt} with random loadings

$$\lambda_{pit} = \lambda_{pi} + \lambda_{pt}$$

where $\lambda_{pt} \sim N(\lambda_p, w_p)$ and $\lambda_{pi} \sim N(0, w_i)$

- The random loadings accommodate measurement non-invariance across time and individual
- Model 4: Adding factor variance non-invariance across time. Can be done either by adding (a) introducing a factor model for the random loadings or (b) introducing a random loadings for the residual of the factor.
- We choose (b). $Var(f) = 0.51 + (0.7 + \sigma_t)^2$ where σ_t is a mean zero random effect

Aggressive-Disruptive Behavior Example Continued: Model 4 Setup

```
model:
%within%
s1-s9 | f by y1-y9;
s | f on time; f@0.51;
ss | e by f; e@1;

%between T1%
y1-y9;
s@0; [s@0];
s1-s9*1; [s1-s9*1];
ss*0.6; [ss@0.7];

%between ID%
y1-y9;
f*1; [f@0];
s*1; [s*0];
s1-s9*1; [s1-s9@0];
ss@0; [ss@0];
```

Aggressive-Disruptive Behavior Example Continued: Results For Categorical Analysis

Between ID Level					
Variances					
Y1	0.113	0.039	0.000	0.045	0.195
Y9	0.194	0.049	0.000	0.108	0.299
S1	0.082	0.029	0.000	0.041	0.152
S9	0.030	0.025	0.000	0.001	0.089
F	1.789	0.241	0.000	1.353	2.300
Between T1 Level					
Means					
S1	0.821	0.094	0.000	0.647	1.018
S9	1.049	0.125	0.000	0.809	1.299
Thresholds					
Y1$1	-0.864	0.156	0.000	-1.183	-0.558
Y2$1	-0.800	0.120	0.000	-1.041	-0.568
Variances					
Y1	0.120	0.231	0.000	0.035	0.666
Y9	0.073	0.158	0.000	0.019	0.416
S1	0.023	0.055	0.000	0.005	0.137
S9	0.050	0.110	0.000	0.011	0.302

- Other extensions of the above model are possible, for example the growth trend can have time specific random effects: f and s can be free over time
- The more clusters there are on a particular level the more elaborate the model can be on that level. However, the more elaborate the model on a particular level is, the slower the convergence
- The main factor f can have a random effect on each of the levels, however the residuals Y_i should be uncorrelated on that level. If they are correlated through another factor model such as, $fb \quad by \quad y_1 - y_9$, then f would be confounded with that factor fb and the model will be poorly identified
- On each level the most general model would be (if there are no random slopes) the unconstrained variance covariance for the dependent variables Y_i. Any model that is a restriction of that model is in principle identified

Aggressive-Disruptive Behavior Example: Conclusions, Continued

- Unlike ML and WLS multivariate modeling, for the time intensive Bayes cross-classified SEM, the more time points there are the more stable and easy to estimate the model is
- Bayesian methods solve problems not feasible with ML or WLS
- Time intensive data naturally fits in the cross-classified modeling framework
- Asparouhov and Muthén (2012). General Random Effect Latent Variable Modeling: Random Subjects, Items, Contexts, and Parameters

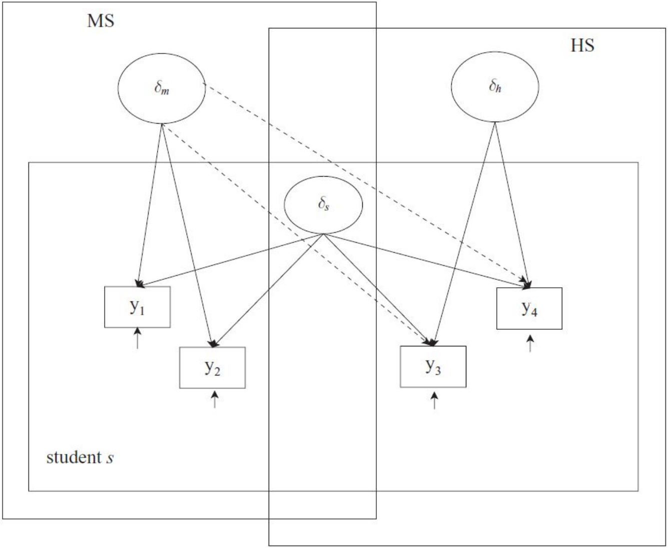
9.6 Cross-Classified / Multiple Membership Applications

- Jeon & Rabe-Hesketh (2012). Profile-Likelihood Approach for Estimating Generalized Linear Mixed Models With Factor Structures. JEBS
- Longitudinal growth model for student self-esteem
- Each student has 4 observations: 2 in middle school in wave 1 and 2, and 2 in high school in wave 3 and 4
- Students have multiple membership: Membership in middle school and in high school with a random effect from both
- Y_{tsmh} is observation at time t for student s in middle school m and high school h

- The model is

$$Y_{tsmh} = \beta_1 + \beta_2 T2 + \beta_3 T3 + \beta_4 T4 + \delta_s + \delta_m \mu_t + \delta_h \lambda_t + \varepsilon_{tsmh}$$

  where T2, T3, T4 are dummy variables for wave 2, 3, 4
- δ_s, δ_m and δ_h are zero mean random effect contributions from student, middle school and high school
- $\mu_t = (1, \mu_2, \mu_3, \mu_4)$
- $\lambda_t = (0, 0, 1, \lambda_4)$, i.e., no contribution from the high school in wave 1 and 2 because the student is still in middle school
- ε_{tsmh} is the residual
- Very simple to setup in Mplus


MS
HS
δ_m
δ_h
δ_s
y_1
y_2
y_3
y_4
student s

```
MODEL:
        %WITHIN%
        fs BY y1-y4@1;
        [y1-y4];

        %BETWEEN mschool%
        fm BY y1@1 y2-y4;
        y1-y4@0; [y1-y4@0];

        %BETWEEN hschool%
        fh BY y1@0 y2@0 y3@1 y4;
        y1-y4@0; [y1-y4@0];
```