

New Developments in Mplus Version 7: Part 2

Bengt Muthén & Tihomir Asparouhov

Mplus
www.statmodel.com

Presentation at Utrecht University
August 2012

Table of Contents I

- 1 Two-Level Random Loadings in IRT: The PISA Data
- 2 Two-Level Random Loadings: Individual Differences Factor Analysis
- 3 Advances in Mixture Modeling: 3-Step Mixture Modeling
- 4 Advances in Mixture Modeling: Latent Transition Analysis Simplifications
 - LTA Example 1: Stage-Sequential Development in Reading Using ECLS-K Data
 - LTA Example 2: Mover-Stayer LTA Modeling of Peer Victimization During Middle School
 - Latent Transition Probabilities: TECH15
 - The LTA Calculator
 - Probability Parameterization
 - Mover-Stayer LTA in Logistic Parameterization
 - Mover-Stayer LTA in Probability Parameterization
- 5 Advances in Mixture Modeling: Exploratory LCA Using Bayesian Analysis

6 3-Level Analysis

- Types of Observed Variables and Random Slopes for 3-Level Analysis
- 3-Level Regression
- 3-Level Regression: Nurses Data
- 3-Level Path Analysis: UG Example 9.21
- 3-Level MIMIC Analysis
- 3-Level Growth Analysis
- TYPE=THREELEVEL COMPLEX
- 3-Level and Cross-Classified Multiple Imputation

7 Cross-Classified Analysis

- Cross-Classified Regression
- Cross-Classified Regression: UG Example 9.24
- Cross-Classified Regression: Pupcross Data
- Cross-Classified Path Analysis: UG Example 9.25

- 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

- 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

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

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

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

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

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

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

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%	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

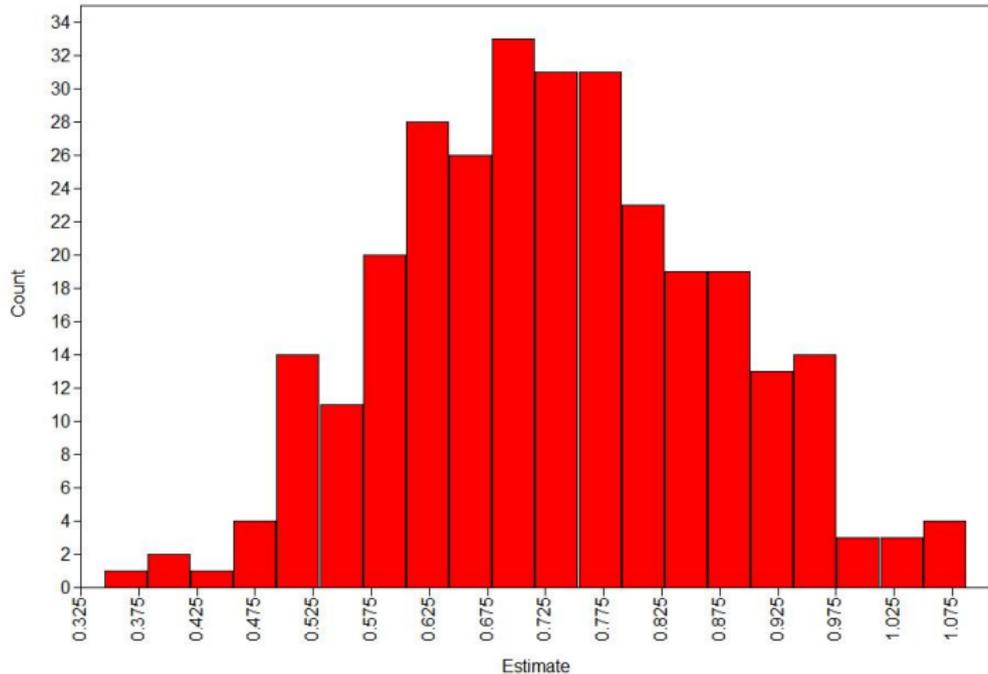
	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I.	
				Lower 2.5%	Upper 2.5%
Between Level					
Variiances					
Y1	0.044	0.017	0.000	0.022	0.087
Y2	0.063	0.025	0.000	0.030	0.124
Y3	0.029	0.011	0.000	0.013	0.057
Y4	0.034	0.012	0.000	0.018	0.066
Y5	0.011	0.006	0.000	0.003	0.027
Y6	0.034	0.016	0.000	0.013	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 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

- 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

- 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 $\text{Var}(f)=0.0000001$. 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 $\text{Var}(f)=0.05$, 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 $\nu < 0.001$, use the prior $\nu \sim \text{IG}(1,0.005)$, and look for the results in the second column. If you want to test the hypothesis $\nu < 0.01$, use the prior $\nu \sim \text{IG}(1,0.05)$, 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

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

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;
```

- 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

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

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

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

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

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

- 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

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 $Var(\epsilon_{jk}) = 0$ 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.

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.

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;
```

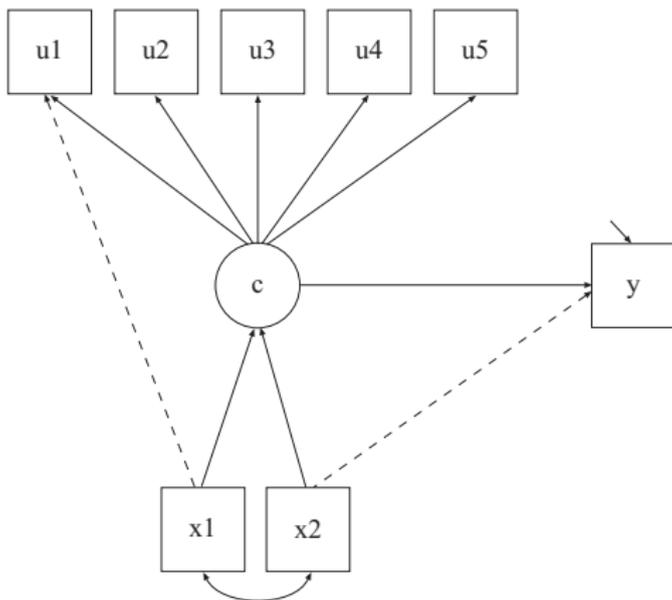
Individual Differences Factor Analysis Results

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

- 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

1-step analysis versus 3-step analysis (analyze-classify-analyze)



However, the one-step approach has certain disadvantages. The first is that it may sometimes be impractical, especially when the number of potential covariates is large, as will typically be the case in a more exploratory study. Each time that a covariate is added or removed not only the prediction model but also the measurement model needs to be reestimated. A second disadvantage is that it introduces additional model building problems, such as whether one should decide about the number of classes in a model with or without covariates. Third, the simultaneous approach does not fit with the logic of most applied researchers, who view introducing covariates as a step that comes after the classification model has been built. Fourth, it assumes that the classification model is built in the same stage of a study as the model used to predict the class membership, which is not necessarily the case.

1-Step vs 3-Step: An Example

Substantive question: Should the latent classes be defined by the indicators alone or also by covariates and distals?

Example: Study of genotypes influencing phenotypes.

Phenotypes may be observed indicators of mental illness such as DSM criteria. The interest is in finding latent classes of subjects and then trying to see if certain genotype variables influence class membership.

Possible objection to 1-step: If the genotypes are part of deciding the latent classes, the assessment of the strength of relationship is compromised.

3-step: Determine the latent classes based on only phenotype information. Then classify subjects. Then relate the classification to the genotypes.

Advantage of 1-Step Over 3-Step

Low-entropy and/or small-sample cases are better handled by 1-step

Direct effects from covariates to latent class indicators and/or distal outcomes may be better handled by 1-step (?)

Prediction from covariates to latent classes and from latent classes to distal outcomes is better handled by 1-step

Auxiliary Variables In Mixture Modeling

- How do we explore a relationship between a latent class variable C and another observed variable X ?
- 1-step ML approach: Estimate the joint model = measurement model (such as LCA) and the C on X model together
- Drawback: Often the latent class formation is affected by X . The latent class variable will not have the intended meaning
- For example, if there is a direct effect from X to an indicator variable the class formation will change if X is included in the model
- How can we estimate the C on X relationship independently of the measurement model?

Auxiliary Variables In Mixture Modeling: The 3-Step Approach

- Prior to Mplus Version 7: Pseudo class (PC) approach. Estimate LCA model, impute C , regress imputed C on X
- New improved method in Mplus Version 7: 3-step approach
 - ① Estimate the LCA model
 - ② Create a nominal most likely class variable N
 - ③ Use a mixture model for N , C and X , where N is a C indicator with measurement error rates prefixed at the uncertainty rate of N estimated in the step 1 LCA analysis
- Mplus Web Note 15. Asparouhov and Muthén (2012). Auxiliary Variables in Mixture Modeling: A 3-Step Approach Using Mplus
- Vermunt (2010). Latent Class Modeling with Covariates: Two Improved Three-Step Approaches. *Political Analysis*, 18, 450-469

- $AUXILIARY=x(R)$; estimates c on x regression using PC method
- $AUXILIARY=y(E)$; estimates y on c , distal outcome model, estimates different mean parameter for y in each class
- $AUXILIARY=x(R3STEP)$; estimates c on x regression using the 3-step method
- $AUXILIARY=y(DU3STEP)$; estimates y on c distal outcome model, assuming unequal variance of y in each class, using the 3-step method
- $AUXILIARY=y(DE3STEP)$; estimates y on c distal outcome model, assuming equal variance of y in each class, using the 3-step method

Auxiliary Variables In Mixture Modeling: Distal Outcomes Simulation

LCA with 2 classes and 5 binary indicators. Mean of Y in class 1 is 0 and in class 2 is 0.7.

Bias/MSE/Coverage

N	Entropy	PC	3-step	1-step
500	0.7	.10/.015/.76	.01/.007/.93	.00/.006/.94
500	0.6	.16/.029/.50	.01/.009/.92	.00/.007/.94
500	0.5	.22/.056/.24	.02/.036/.80	.01/.012/.96
2000	0.7	.10/.011/.23	.02/.002/.91	.00/.002/.93
2000	0.6	.15/.025/.03	.03/.003/.86	.00/.002/.94
2000	0.5	.22/.051/.00	.00/.015/.76	.00/.003/.94

Auxiliary Variables In Mixture Modeling: Distal Outcomes Simulation, Continued

- The 3-step procedure outperforms the PC procedure substantially in terms of bias, mean squared error and confidence interval coverage
- The 3-step procedure procedure loss of efficiency is not substantial when compared to the 1-step procedure when the class separation is not poor (entropy of 0.6 or higher)
- The 3-step procedure with unequal variance is better than the 3-step procedure with equal variance when the equal variance assumption doesn't hold.

Auxiliary Variables In Mixture Modeling: Latent Class Predictor Simulation

LCA with 2 classes and 5 binary indicators. C on X is 0.5.

Bias/MSE/Coverage

N	Entropy	PC	3-step	1-step
500	0.7	.13/.023/.84	.01/.015/.96	.01/.014/.95
500	0.6	.20/.044/.59	.00/.020/.95	.01/.017/.96
500	0.5	.28/.083/.24	.01/.035/.95	.03/.028/.97
2000	0.7	.13/.019/.24	.00/.004/.93	.00/.004/.94
2000	0.6	.20/.042/.01	.00/.004/.96	.00/.004/.94
2000	0.5	.29/.085/.00	.01/.010/.93	.01/.006/.95

Auxiliary Variables In Mixture Modeling: Latent Class Predictor Simulation, Continued

- The 3-step procedure again outperforms the PC procedure substantially in terms of bias, mean squared error and confidence interval coverage
- The efficiency of the 3-step procedure is comparable to the 1-step method when the entropy is 0.6 or higher

- Clark & Muthén (2009). Relating latent class analysis results to variables not included in the analysis
- Extensive simulation of other methods, such as Most Likely Class Membership Regression, Probability Regression, Probability-Weighted Regression, even Bayes Imputations, show that all these methods have biased estimates

Auxiliary Variables In Mixture Modeling: Latent Class Predictor Example

```
VARIABLE:  NAMES = u1-u5 x;  
           CATEGORICAL = u1-u5;  
           CLASSES = c(2);  
           AUXILIARY = x(R3STEP);  
DATA:      FILE = 1.dat;  
ANALYSIS:  TYPE = MIXTURE;  
MODEL:     !no model is needed, LCA is default
```

Auxiliary Variables In Mixture Modeling: Latent Class Predictor Example

TESTS OF CATEGORICAL LATENT VARIABLE MULTINOMIAL LOGISTIC REGRESSIONS
THE 3-STEP PROCEDURE

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
C#1	ON				
X		0.488	0.190	2.569	0.010

Auxiliary Variables In Mixture Modeling

- The latent class variable can be identified by any mixture model, not just LCA, for example Growth Mixture Models
- Multiple auxiliary variables can be analyzed at the same time
- Auxiliary variables can be included in a Montecarlo setup
- The 3-step procedure can be setup manually for other types of models, different from the distal outcome model and the latent class regression. For example, distal outcomes regressed on the latent class variable and another predictor

3-Step Mixture Modeling For Special Models

- How can we estimate a mixture regression model independently of the LCA model that defines C

$$Y = \alpha_c + \beta_c X + \varepsilon$$

- We simulate data with $\alpha_1 = 0$, $\alpha_2 = 1$, $\beta_1 = 0.5$, $\beta_2 = -0.5$
- Step 1: Estimate the LCA model (without the auxiliary model) with the following option
SAVEDATA: FILE=1.dat; SAVE=CPROB;
- The above option creates the most likely class variable N
- Step 2: Compute the error rate for N . In the LCA output find

Average Latent Class Probabilities for Most Likely Latent Class Membership
by Latent Class (Column)

	1	2
1	0.835	0.165
2	0.105	0.895

- Compute the nominal variable N parameters

$$\log(0.835/0.165) = 1.621486$$

$$\log(0.105/0.895) = -2.14286$$

- Step 3: estimate the model where N is a latent class indicator with the above fixed parameters and include the class specific Y on X model
- When the class separation in the LCA is pretty good then N is almost a perfect C indicator

```
VARIABLE:  NAMES = u1-u5 y x p1 p2 n;  
           NOMINAL = n;  
           CLASSES = c(2);  
           USEVARIABLES = y x n;
```

```
MODEL:
```

```
  %OVERALL%
```

```
  y ON x;
```

```
  %c#1%
```

```
  [n#1@1.621486];
```

```
  y ON x;
```

```
  %c#2%
```

```
  [n#1@-2.14286];
```

```
  y ON x;
```

3-Step Mixture Modeling For Special Models Final Results

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Latent Class 1				
Y				
X	0.546	0.054	10.185	0.000
ON				
Means				
N#1	1.621	0.000	999.000	999.000
Intercepts				
Y	0.101	0.052	1.961	0.050
Latent Class 2				
Y				
X	-0.483	0.051	-9.496	0.000
ON				
Means				
N#1	-2.143	0.000	999.000	999.000
Intercepts				
Y	1.037	0.056	18.498	0.000

New developments in Version 7:

- TECH15 output with conditional class probabilities useful for studying transition probabilities with an observed binary covariate such as treatment/control or a latent class covariate
- LTA transition probability calculator for continuous covariates
- Probability parameterization to simplify input for Mover-Stayer LTA and other models with restrictions on the transition probabilities

LTA Example 1: Stage-Sequential Development in Reading Using ECLS-K Data

Kaplan (2008). An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology*, 44, 457-467.

- Early Childhood Longitudinal Study - Kindergarten cohort
- Four time points: Kindergarten Fall, Spring and Grade 1 Fall, Spring; $n = 3,575$
- Five dichotomous proficiency scores: Letter recognition, beginning sounds, ending letter sounds, sight words, words in context
- Binary poverty index
- LCA suggests 3 classes: Low alphabet knowledge (LAK), early word reading (EWR), and early reading comprehension (ERC)

- Three latent classes:
 - Class 1: Low alphabet knowledge (LAK)
 - Class 2: Early word reading (EWR)
 - Class 3: Early reading comprehension (ERC)

The ECLS-K LTA model has the special feature of specifying no decline in knowledge as zero transition probabilities. For example, transition from Kindergarten Fall to Spring:

LATENT TRANSITION PROBABILITIES
BASED ON THE ESTIMATED MODEL

c1 classes (rows) by c2 classes (columns)

	1	2	3
1	0.329	0.655	0.017
2	0.000	0.646	0.354
3	0.000	0.000	1.000

LTA Example 1: ECLS-K.

Transition Tables for the Binary Covariate Poverty using TECH15

		Poverty = 0 (cp=1)			Poverty = 1 (cp=2)		
		c2			c2		
		1	2	3	1	2	3
c1	1	0.252	0.732	0.017	0.545	0.442	0.013
	2	0.000	0.647	0.353	0.000	0.620	0.380
	3	0.000	0.000	1.000	0.000	0.000	1.000

LTA Example 2: Mover-Stayer LTA Modeling of Peer Victimization During Middle School

Nylund (2007) Doctoral dissertation: Latent Transition Analysis: Modeling Extensions and an Application to Peer Victimization

- Student's self-reported peer victimization in Grade 6, 7, and 8
- Low SES, ethnically diverse public middle schools in the Los Angeles area (11% Caucasian, 17% Black, 48 % Latino, 12% Asian)
- $n = 2045$
- 6 binary items: Picked on, laughed at, called bad names, hit and pushed around, gossiped about, things taken or messed up (Neary & Joseph, 1994 Peer Victimization Scale)

LTA Example 2: Mover-Stayer Model

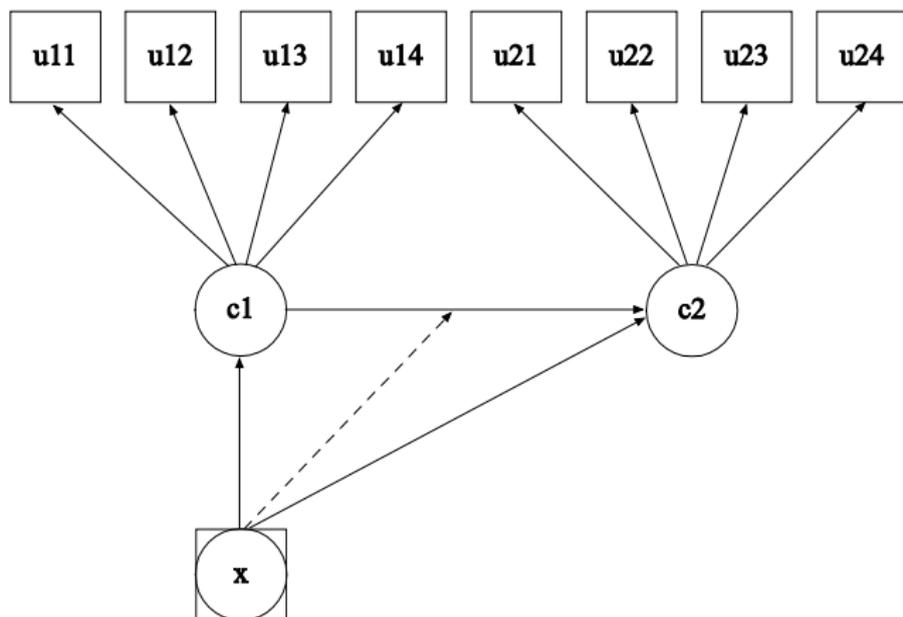
- Class 1: Victimized (G6-G8: 19%, 10%, 8%)
- Class 2: Sometimes victimized (G6-G8: 34%, 27%, 21%)
- Class 3: Non-victimized (G6-G8: 47%, 63%, 71%)

Movers (60%)									
		c2 (Grade 7)					c3 (Grade 8)		
c1 (Grade 6)		0.29	0.45	0.26	c2 (Grade 7)		0.23	0.59	0.18
		0.06	0.44	0.51			0.04	0.47	0.49
		0.04	0.46	0.55			0.06	0.17	0.77

Stayers (40%)									
		c2 (Grade 7)					c3 (Grade 8)		
c1 (Grade 6)		1	0	0	c2 (Grade 7)		1	0	0
		0	1	0			0	1	0
		0	0	1			0	0	1

Latent Transition Probabilities Influenced by a Categorical Covariate: Using TECH15

- Covariate treated as Knownclass
- New feature in Version 7: TECH 15 output - conditional probabilities for "c ON c"



```
TITLE:      this is an example of a LTA with a
             binary covariate representing group
DATA:       FILE = ex8.13.dat;
VARIABLE:   NAMES = u11-u15 u21-u25 g;
             CATEGORICAL = u11-u15 u21-u25;
             CLASSES = cg(2) c1(3) c2(3);
             KNOWNCLASS = cg(g=0 g=1);
ANALYSIS:   TYPE=MIXTURE;
MODEL:      %OVERALL%
             c1 c2 ON cg;
MODEL cg:   %cg#1%
             c2 ON c1;
             %cg#2%
             c2 ON c1;
MODEL u11:  g=1#10
```

LTA with a Binary Covariate as Knownclass: UG Ex8.13, Continued

```
MODEL c1: %c1#1%  
          [u11$1] (1);  
          [u12$1] (2);  
          [u13$1] (3);  
          [u14$1] (4);  
          [u15$1] (5);  
          %c1#2%  
          [u11$1] (6);  
          [u12$1] (7);  
          [u13$1] (8);  
          [u14$1] (9);  
          [u15$1] (10);  
          %c1#3%  
          [u11$1] (11);  
          [u12$1] (12);  
          [u13$1] (13);  
          [u14$1] (14);  
          [u15$1] (15);
```

LTA with a Binary Covariate as Knownclass: UG Ex8.13, Continued

```
MODEL c2:
      %c2#1%
      [u21$1] (1);
      [u22$1] (2);
      [u23$1] (3);
      [u24$1] (4);
      [u25$1] (5);
      %c2#2%
      [u21$1] (6);
      [u22$1] (7);
      [u23$1] (8);
      [u24$1] (9);
      [u25$1] (10);
      %c2#3%
      [u21$1] (11);
      [u22$1] (12);
      [u23$1] (13);
      [u24$1] (14);
      [u25$1] (15);
OUTPUT:  TECH1 TECH8 TECH15;
```

Consider the logit parameterization for CLASSES = c1(3) c2(3):

		c2		
		1	2	3
c1	1	$a_1 + b_{11}$	$a_2 + b_{21}$	0
	2	$a_1 + b_{12}$	$a_2 + b_{22}$	0
	3	a_1	a_2	0

where each row shows the coefficients for a multinomial logistic regression of c2 on c1 with the last c2 class as reference class.

Zero lower-triangular probabilities are obtained by fixing the a_1 , a_2 , and b_{12} parameters at the logit value -15. The parameters b_{11} , b_{21} , and b_{22} are estimated.

LTA Example 1: ECLS-K, Mplus Input

```
TITLE:          LTA of Kindergarten Fall and Spring (3 x 3)
DATA:          FILE = dp.analytic.dat;
              FORMAT = f1.0, 20f2.0;
VARIABLE:     NAMES = pov letrec1 begin1 ending1 sight1 wic1
              letrec2 begin2 ending2 sight2 wic2
              letrec3 begin3 ending3 sight3 wic3
              letrec4 begin4 ending4 sight4 wic4;
              USEVARIABLES = letrec1 begin1 ending1 sight1 wic1
              letrec2 begin2 ending2 sight2 wic2;
              ! letrec3 begin3 ending3 sight3 wic3
              ! letrec4 begin4 ending4 sight4 wic4;
              CATEGORICAL = letrec1 begin1 ending1 sight1 wic1
              letrec2 begin2 ending2 sight2 wic2;
              ! letrec3 begin3 ending3 sight3 wic3
              ! letrec4 begin4 ending4 sight4 wic4;
              CLASSES = c1(3) c2(3);
              MISSING = .;
```

```
ANALYSIS: TYPE = MIXTURE;  
          STARTS = 400 80;  
          PROCESSORS = 8;  
MODEL:   %OVERALL%  
          ! fix lower triangular transition probabilities = 0:  
          [c2#1@-15 c2#2@-15]; ! fix a1 = a2 = -15  
          c2#1 ON c1#2@-15; ! fix b12 = -15  
          c2#1 ON c1#1*15; ! b11: start at 15 to make total logit start=0  
          c2#2 ON c1#1-c1#2*15; ! b21, b22
```

```
MODEL c1:    %c1#1%  
             [letrec1$1-wic1$1] (1-5) ;  
             %c1#2%  
             [letrec1$1-wic1$1] (6-10);  
             %c1#3%  
             [letrec1$1-wic1$1] (11-15);  
MODEL c2:    %c2#1%  
             [letrec2$1-wic2$1] (1-5);  
             %c2#2%  
             [letrec2$1-wic2$1] (6-10);  
             %c2#3%  
             [letrec2$1-wic2$1] (11-15);  
OUTPUT:     TECH1 TECH15;  
PLOT:       TYPE = PLOT3;  
            SERIES = letrec1-wic1(*) | letrec2-wic2(*);
```

Output: Latent Transition Table

LATENT TRANSITION PROBABILITIES BASED ON THE ESTIMATED MODEL

c1 classes (rows) by c2 classes (columns)

	1	2	3
1	0.329	0.655	0.017
2	0.000	0.646	0.354
3	0.000	0.000	1.000

LTA Example 1: ECLS-K, Mplus Input, Adding Poverty as Knownclass

```
CLASSES = cp(2) c1(3) c2(3);
KNOWNCLASS = cp(pov=0 pov=1);
ANALYSIS: TYPE = MIXTURE;
STARTS = 400 80; PROCESSORS = 8;
MODEL: %OVERALL%
c1 ON cp;
[c2#1@-15 c2#2@-15];
c2#1 ON c1#2@-15;
MODEL cp: %cp#1%
c2#1 ON c1#1*15;
c2#2 ON c1#1-c1#2*15;
%cp#2%
c2#1 ON c1#1*15;
c2#2 ON c1#1-c1#2*15;
MODEL c1: %c1#1% etc as before
```

$$P(CP=1)=0.808$$

$$P(CP=2)=0.192$$

$$P(C1=1|CP=1)=0.617$$

$$P(C1=2|CP=1)=0.351$$

$$P(C1=3|CP=1)=0.032$$

$$P(C1=1|CP=2)=0.872$$

$$P(C1=2|CP=2)=0.123$$

$$P(C1=3|CP=2)=0.005$$

$$P(C2=1|CP=1,C1=1)=0.252$$

$$P(C2=2|CP=1,C1=1)=0.732$$

$$P(C2=3|CP=1,C1=1)=0.017$$

$$P(C2=1|CP=1,C1=2)=0.000$$

$$P(C2=2|CP=1,C1=2)=0.647$$

$$P(C2=3|CP=1,C1=2)=0.353$$

$$P(C2=1 | CP=1, C1=3)=0.000$$

$$P(C2=2 | CP=1, C1=3)=0.000$$

$$P(C2=3 | CP=1, C1=3)=1.000$$

$$P(C2=1 | CP=2, C1=1)=0.545$$

$$P(C2=2 | CP=2, C1=1)=0.442$$

$$P(C2=3 | CP=2, C1=1)=0.013$$

$$P(C2=1 | CP=2, C1=2)=0.000$$

$$P(C2=2 | CP=2, C1=2)=0.620$$

$$P(C2=3 | CP=2, C1=2)=0.380$$

$$P(C2=1 | CP=2, C1=3)=0.000$$

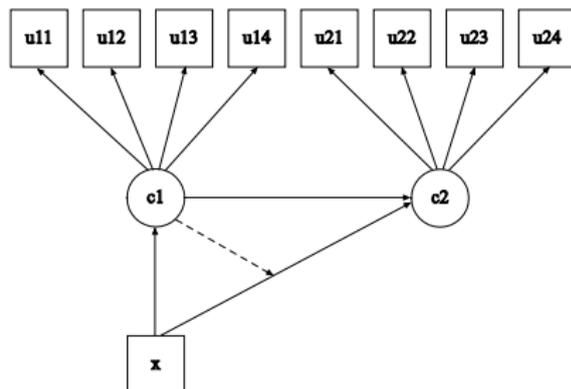
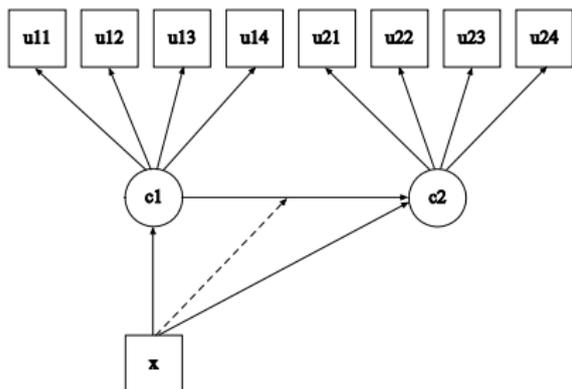
$$P(C2=2 | CP=2, C1=3)=0.000$$

$$P(C2=3 | CP=2, C1=3)=1.000$$

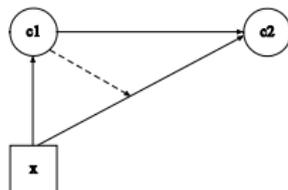
Latent Transition Probabilities Influenced by a Continuous Covariate

- Muthén & Asparouhov (2011). LTA in Mplus: Transition probabilities influenced by covariates. Mplus Web Notes: No. 13. July 27, 2011. www.statmodel.com
- New feature in Version 7: The LTA calculator

Interaction Displayed Two Equivalent Ways



Review of Logit Parameterization with Covariates: Parameterization 2



MODEL: %OVERALL%
 c1 ON x;
 c2 ON c1;

MODEL c1: %c1#1%
 c2#1 ON x (g11);
 c2#2 ON x (g21);

 %c1#2%
 c2#1 ON x (g12);
 c2#2 ON x (g22);

 %c1#3%
 c2#1 ON x (g13);
 c2#2 ON x (g23);

		c2		
		1	2	3
c1	1	$a_1 + b_{11} + g_{11} x$	$a_2 + b_{21} + g_{21} x$	0
	2	$a_1 + b_{12} + g_{12} x$	$a_2 + b_{22} + g_{22} x$	0
	3	$a_1 + g_{13} x$	$a_2 + g_{23} x$	0

LTA Example 1: ECLS-K, Adding Poverty as Covariate

```
USEVARIABLES = letrec1 begin1 ending1 sight1 wic1  
letrec2 begin2 ending2 sight2 wic2 pov;  
! letrec3 begin3 ending3 sight3 wic3  
! letrec4 begin4 ending4 sight4 wic4;  
CATEGORICAL = letrec1 begin1 ending1 sight1 wic1  
letrec2 begin2 ending2 sight2 wic2;  
! letrec3 begin3 ending3 sight3 wic3  
! letrec4 begin4 ending4 sight4 wic4;  
CLASSES = c1(3) c2(3);  
MISSING = .;  
ANALYSIS: TYPE = MIXTURE;  
STARTS = 400 80;  
PROCESSORS = 8;
```

LTA Example 1: ECLS-K, Adding Poverty as Covariate, Continued

MODEL: %OVERALL%
 c1 ON pov;
 ! do c2 ON pov in c1-specific model part to get interaction
 [c2#1@-15 c2#2@-15]; ! to give zero probability of declining
 c2#1 ON c1#2@-15; ! to give zero probability of declining
 c2#1 ON c1#1*15;
 c2#2 ON c1#1-c1#2*15;

MODEL c1: %c1#1%
 c2 ON pov; ! (g11) and (g21)
 %c1#2%
 c2#1 ON pov@-15; ! to give zero probability of declining (g12)
 c2#2 ON pov; ! (g22)

 ! %c1#3% not mentioned due to g13=0, g23=0 by default

LTA Calculator Applied to Poverty

The screenshot displays the Mplus software interface. The main window title is "Mplus - [8-14 final lta first and second tech15 poverty as x 6-23.out]". The menu bar includes File, Edit, View, Mplus, Plot, Window, and Help. The "Mplus" menu is open, showing options: Language Generator, Run Mplus (Alt+R), and LTA calculator. The LTA calculator option is highlighted.

The "Covariate values for adjusted estimated means" dialog box is open. It contains the following elements:

- Section: "Covariate sets for adjusted estimated means"
- Text: "Choose one of the following: name a new covariate set, display a covariate set, remove a covariate set."
- Buttons: "Name covariate set", "Display covariate set", "Remove covariate set"
- Fields: "pov" (selected in both "Display covariate set" and "Remove covariate set" dropdowns)
- Text: "Number of graphs for each group: 1"
- Section: "Covariate values"
- Table:

Variable	Value
POV	
- Section: "Set value" (sub-dialog box)
 - Text: "Value setting for: POV"
 - Radio buttons: "Use sample mean" (selected), "Use 0" (with field "0")
 - Buttons: "OK", "Cancel"
- Text: "Use sample mean" (button)
- Text: "Set" (button)
- Text: "Use sample mean" (button)
- Buttons: "OK", "Cancel" (at the bottom)

Estimated conditional probabilities for the latent class variables:

Condition(s): $POV = 1.000000$

$$P(C1=1)=0.872$$

$$P(C1=2)=0.123$$

$$P(C1=3)=0.005$$

$$P(C2=1|C1=1)=0.545$$

$$P(C2=2|C1=1)=0.442$$

$$P(C2=3|C1=1)=0.013$$

$$P(C2=1|C1=2)=0.000$$

$$P(C2=2|C1=2)=0.620$$

$$P(C2=3|C1=2)=0.380$$

$$P(C2=1|C1=3)=0.000$$

$$P(C2=2|C1=3)=0.000$$

$$P(C2=3|C1=3)=1.000$$

Latent Transition Analysis: Probability Parameterization

LTA models that do not have continuous x 's can be more conveniently specified using `PARAMETERIZATION=PROBABILITY`. Consider the probability parameterization for `CLASSES = c1(3) c2(3)`:

		c2		
		1	2	3
c1	1	p11	p12	0
	2	p21	p22	0
	3	p31	p32	0

where the probabilities in each row add to 1 and the last c2 class is not mentioned. The p parameters are referred to using `ON`. The latent class variable c1 which is the predictor has probability parameters [c1#1 c1#2], whereas "intercept" parameters are not included for c2.

A transition probability can be conveniently fixed at 1 or 0 by using the p parameters.

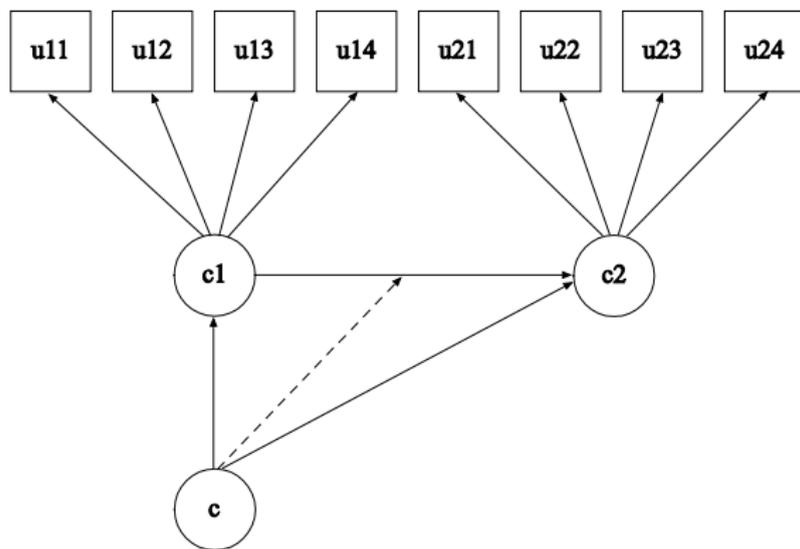
LTA Example 1: ECLS-K, Mplus Input Using Probability Parameterization

```
CLASSES = c1(3) c2(3);
```

```
ANALYSIS:  TYPE = MIXTURE;  
           STARTS = 400 80;  
           PROCESSORS = 8;  
           PARAMETERIZATION=PROBABILITY;
```

```
MODEL:    %OVERALL%  
          c2 ON c1; ! frees 6 transition probabilities  
          ! fix lower triangular transition probabilities = 0:  
          c2#1 ON c1#2-c1#3@0;  
          c2#2 ON c1#3@0;
```

Mover-Stayer LTA in Logistic Parameterization



FAQ: 9/23/11: LTA with Movers-Stayers (www.statmodel.com)

This uses the regular logit parameterization which is complex for Mover-Stayer modeling.

Mover-Stayer LTA in Logistic Parameterization: No Covariates (Nylund example)

- $c(2)$ is the mover-stayer latent class variable:
 $c\#1$ is the mover class, $c\#2$ is the stayer class
- $c1(3)$, $c2(3)$, $c3(3)$ are time-specific latent class variables

MODEL: %OVERALL%
 ! Relating $c1$, $c2$, $c3$ to c : (Movers)
 $c1\ c2\ c3\ ON\ c$;

```
! Relating c2 and c3 to c: (Stayers)
[c2#1@-15]; !"a1" - probability of transitioning from problem
!at time 1 to problem-free at time 2 is fixed
!at zero for the stayer class
[c2#2@-15];
[c3#1@-15 c3#2@-15];
```

```
!covariates are not used in this analysis:
! c1#1 c1#2 ON female schsafe2 cdi2 sas2;
! c2#1 c2#2 ON female schsafe4 cdi4 sas4;
! c3#1 c3#2 ON female schsafe6 cdi6 sas6;
! c#1 ON female afam latino asian birace;
```

MODEL c:

%c#1% !mover class

c2 ON c1;

c3 ON c2;

%c#2% !stayer class

c2#1 ON c1#1@30; !"b11" - stayer class has prob 1 of staying

c2#1 ON c1#2@-45; !b12

c2#2 ON c1#1@-45; !b21

c2#2 ON c1#2@30; !b22

c3#1 ON c2#1@30; !"b11" - stayer class has prob 1 of staying

c3#1 ON c2#2@-45; !b12

c3#2 ON c2#1@-45; !b21

c3#2 ON c2#2@30; !b22

Measurement Part of the Model

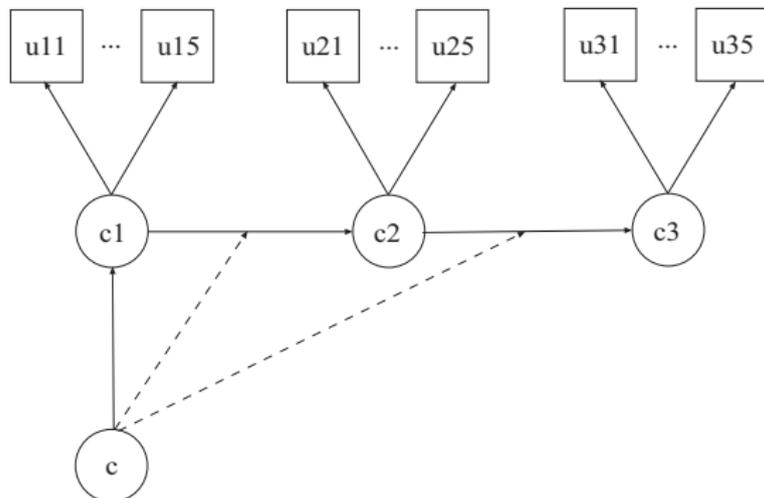
MODEL c.c1: %c#1.c1#1% !mover
 [bpvs1b\$1- bpvs6b\$1*-5] (1-6);
 %c#1.c1#2%
 [bpvs1b\$1- bpvs6b\$1*1] (7-12);
 %c#1.c1#3%
 [bpvs1b\$1- bpvs6b\$1*4] (13-18);
 %c#2.c1#1% !stayer
 [bpvs1b\$1- bpvs6b\$1*-5] (1-6);
 %c#2.c1#2%
 [bpvs1b\$1- bpvs6b\$1*1] (7-12);
 %c#2.c1#3%
 [bpvs1b\$1- bpvs6b\$1*4] (13-18);

MODEL c.c2: %c#1.c2#1%
 [bpvs1d\$1- bpvs6d\$1*-5] (1-6);
 %c#1.c2#2%
 [bpvs1d\$1- bpvs6d\$1*1] (7-12);
 %c#1.c2#3%
 [bpvs1d\$1- bpvs6d\$1*4] (13-18);

MODEL c.c3:

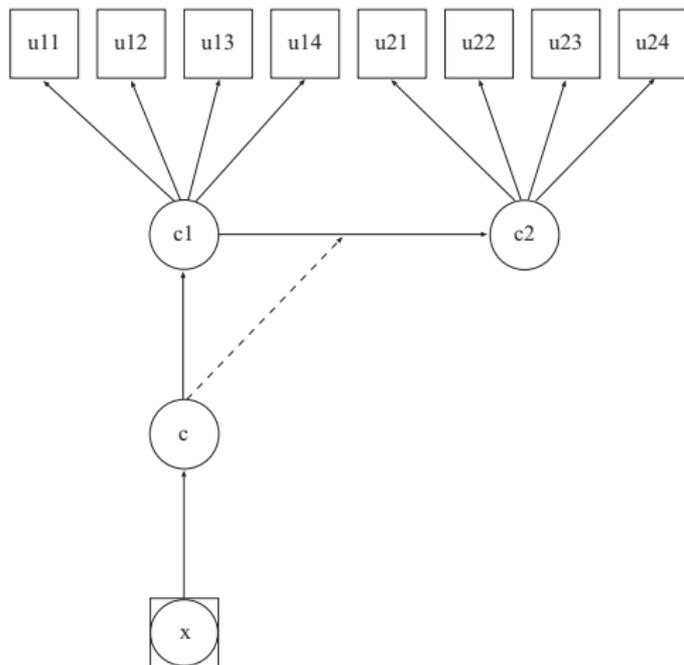
%c#2.c2#1%
[bpvs1d\$1- bpvs6d\$1*-5] (1-6);
%c#2.c2#2%
[bpvs1d\$1- bpvs6d\$1*1] (7-12);
%c#2.c2#3%
[bpvs1d\$1- bpvs6d\$1*4] (13-18);
%c#1.c3#1%
[bpvs1f\$1- bpvs6f\$1*-5] (1-6);
%c#1.c3#2%
[bpvs1f\$1- bpvs6f\$1*1] (7-12);
%c#1.c3#3%
[bpvs1f\$1- bpvs6f\$1*4] (13-18);
%c#2.c3#1%
[bpvs1f\$1- bpvs6f\$1*-5] (1-6);
%c#2.c3#2%
[bpvs1f\$1- bpvs6f\$1*1] (7-12);
%c#2.c3#3%
[bpvs1f\$1- bpvs6f\$1*4] (13-18);

Mover-Stayer LTA in Probability Parameterization



```
PARAMETERIZATION = PROBABILITY;  
MODEL: %OVERALL% ! Relating c1 to c:  
c1 ON c;  
MODEL c: %c#1% ! Mover class  
c2 ON c1;  
c3 ON c2;  
%c#2% ! Stayer class  
c2#1 ON c1#1@1; c2#2 ON c1#1@0;  
c2#1 ON c1#2@0; c2#2 ON c1#2@1;  
c2#1 ON c1#3@0; c2#2 ON c1#3@0;  
  
c3#1 ON c2#1@1; c3#2 ON c2#1@0;  
c3#1 ON c2#2@0; c3#2 ON c2#2@1;  
c3#1 ON c2#3@0; c3#2 ON c2#3@0;  
!measurement part as before
```

Mover-Stayer LTA in Probability Parameterization: Predicting Mover-Stayer Class Membership from a Nominal Covariate



```

CLASSES = cg(5) c (2) c1(3) c2(3) c3(3);
KNOWNCLASS = cg(eth=0 eth=1 eth=2 eth=3 eth=4);
ANALYSIS: TYPE = MIXTURE COMPLEX;
STARTS = 400 100;
PROCESS = 8;
PARAMETERIZATION = PROBABILITY;
MODEL: %OVERALL%
c ON cg#1-cg#5 (b1-b5);
c1 ON c;
MODEL c: etc
MODEL CONSTRAINT:
NEW(logor2-logor5);
! log of ratio of odds of being Mover vs Stayer for the groups
logor2 = log((b2/(1-b2))/(b1/(1-b1))); ! eth=1 (cg=2) vs 0 (cg=1)
logor3 = log((b3/(1-b3))/(b1/(1-b1))); ! eth=2 (cg=3) vs 0
logor4 = log((b4/(1-b4))/(b1/(1-b1))); ! eth=3 (cg=4) vs 0
logor5 = log((b5/(1-b5))/(b1/(1-b1))); ! eth=4 (cg=5) vs 0

```

Asparouhov, T. & Muthén, B. (2011). Using Bayesian priors for more flexible latent class analysis. Proceedings of the 2011 Joint Statistical Meetings.

- Let Y_i be binary observed variable (works only with binary indicators), C latent class variable.
- Conditional independence LCA model

$$P(Y_1, \dots, Y_m | C) = P(Y_1 | C) \dots P(Y_m | C) = p_{1c} \dots p_{mc}$$

- LCA model without conditional independence using probit link

$$P(Y_i = 1 | C) = P(Y_i^* > 0 | C)$$

$$Y_i^* | C \sim N(\mu_{ic}, 1)$$

- The joint distribution of Y_i

$$Y_1^*, \dots, Y_m^* | C \sim N(\mu_c, \Sigma_c)$$

where Σ_c is a correlation matrix.

- If $\Sigma_c = I$ the model is conditionally independent

- Specify that within class the items are approximately independent

$\Sigma \sim$ marginal correlation distribution of $IW(I, DF)$.

- By increasing DF the prior forces more independence between the indicators, decreasing DF allows for more dependence between the indicators, DF are explored as a sensitivity analysis
- This methodology formalizes the belief that conditional independence is only approximate
- Automatically and easily detects residual correlations among indicators, more flexible LCA
- The "tiny" prior leads to more stable estimation than unconstrained correlations with uninformative priors
- Can easily model any number of (full or partial) correlations (compared to ML which yields 1 dimension of integration for every correlation)

Why Do We Need To Account For Residual Correlation Between Indicators?

- The number of classes can be overestimated.
- Example: 2-class, 6 binary indicators, $\mu = \pm 1$, $n=5000$
- In class 1: $Corr(U_1, U_2) = 0.8$

Number of Classes	Assumption	BIC
2	independence	32301
3	independence	31893
4	independence	31945
2	correlation	31858
3	correlation	31901

Qu-Tan-Kutner Example

- Qu T, Tan M, Kutner MH. Random-effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 52:797-810, 1996.
- 4 binary indicators, 2 class LCA model, correlation between indicators 2 and 3 in class 2 using a factor variable
- LCA with unconstrained Σ and $IW(I, 15)$ (unidentified model) finds only one significant correlation: indicators 2 and 3 in class 2.
- Bayes LCA with one correlation yields the same results as Qu-Kutner ML analysis.

```
DATA:                               FILE = AlvordHIV1.dat;
VARIABLE:                            NAMES = u1-u4;
                                      CATEGORICAL = u1-u4;
                                      CLASSES = c(2);
ANALYSIS:                             TYPE = MIXTURE;
                                      ESTIMATOR = BAYES;
                                      FBITERATIONS = 10000;
MODEL:                                %OVERALL%
                                      %c#1%
                                      u1-u4 WITH u1-u4 (p1-p6);
                                      %c#2%
                                      u1-u4 WITH u1-u4 (q1-q6);
MODEL PRIORS:                         p1-p6~IW(0, 15);
                                      pq-q6~IW(0, 15);
```

Qu-Kutner Example: Bayes Exploratory LCA

	Estimate	Posterior S.D.	One-Tailed P-Value	95% Lower 2.5%	C.I. Upper 2.5%
Latent Class 1					
U1 WITH					
U2	0.138	0.186	0.235	-0.220	0.492
U3	0.000	0.217	0.500	-0.423	0.417
U4	-0.016	0.220	0.474	-0.461	0.384
U2 WITH					
U3	-0.008	0.222	0.485	-0.422	0.428
U4	-0.105	0.196	0.299	-0.479	0.283
U3 WITH					
U4	-0.004	0.215	0.493	-0.414	0.411
Latent Class 2					
U1 WITH					
U2	0.007	0.225	0.487	-0.440	0.428
U3	0.001	0.219	0.499	-0.422	0.429
U4	-0.018	0.206	0.465	-0.393	0.395
U2 WITH					
U3	0.352	0.108	0.001	0.129	0.546
U4	0.011	0.216	0.480	-0.414	0.417
U3 WITH					
U4	0.008	0.218	0.487	-0.406	0.420

Continuous outcomes: ML and Bayesian estimation

Categorical outcomes: Bayesian estimation (Bayes uses probit)

Count and nominal outcomes: Not yet available

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

- Three level multivariate data Y_{pijk}

$$Y_{pijk} = Y_{1pijk} + Y_{2pj k} + Y_{3pk} \quad (1)$$

- 3 sets of structural equations - one on each level

$$Y_{1ijk} = \Lambda_1 \eta_{ijk} + \varepsilon_{ijk} \quad (2)$$

$$\eta_{ijk} = B_1 \eta_{ijk} + \Gamma_1 x_{ijk} + \xi_{ijk}. \quad (3)$$

$$Y_{2jk} = \Lambda_2 \eta_{jk} + \varepsilon_{jk} \quad (4)$$

$$\eta_{jk} = B_2 \eta_{jk} + \Gamma_2 x_{jk} + \xi_{jk}. \quad (5)$$

$$Y_{3k} = \nu + \Lambda_3 \eta_k + \varepsilon_k \quad (6)$$

$$\eta_k = \alpha + B_3 \eta_k + \Gamma_3 x_k + \xi_k. \quad (7)$$

- Type 1: Defined on the level 1
%WITHIN%
 $s \mid y \text{ ON } x;$
The random slope s has variance on level 2 and level 3
- Type 2: Defined on the level 2
%BETWEEN level2%
 $s \mid y \text{ ON } x;$
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

3-Level Regression

$$\text{Level 1 : } y_{ijk} = \beta_{0jk} + \beta_{1jk} x_{ijk} + \varepsilon_{ijk}, \quad (1)$$

$$\text{Level 2a : } \beta_{0jk} = \gamma_{00k} + \gamma_{01k} w_{jk} + \zeta_{0jk}, \quad (2)$$

$$\text{Level 2b : } \beta_{1jk} = \gamma_{10k} + \gamma_{11k} w_{jk} + \zeta_{1jk}, \quad (3)$$

$$\text{Level 3a : } \gamma_{00k} = \kappa_{000} + \kappa_{001} z_k + \delta_{00k}, \quad (4)$$

$$\text{Level 3b : } \gamma_{01k} = \kappa_{010} + \kappa_{011} z_k + \delta_{01k}, \quad (5)$$

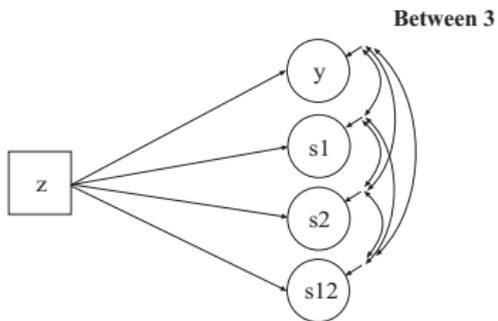
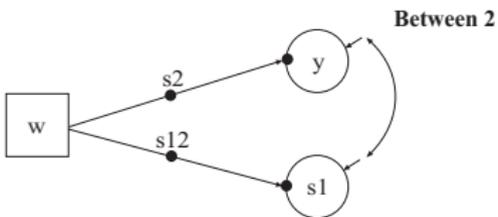
$$\text{Level 3c : } \gamma_{10k} = \kappa_{100} + \kappa_{101} z_k + \delta_{10k}, \quad (6)$$

$$\text{Level 3d : } \gamma_{11k} = \kappa_{110} + \kappa_{111} z_k + \delta_{11k}, \quad (7)$$

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

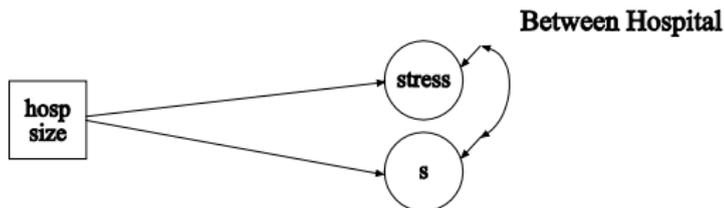
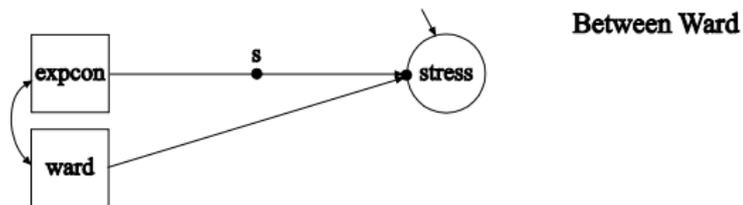
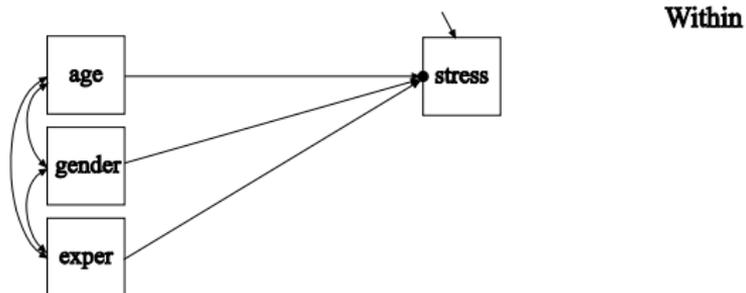
```
TITLE:      this is an example of a three-level
            regression with a continuous dependent
            variable
DATA:       FILE = ex9.20.dat;
VARIABLE:   NAMES = y x w z level2 level3;
            CLUSTER = level3 level2;
            WITHIN = x;
            BETWEEN =(level2) w (level3) z;
ANALYSIS:   TYPE = THREELLEVEL 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;
```

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
zexpcn cexpcn 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;

```
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

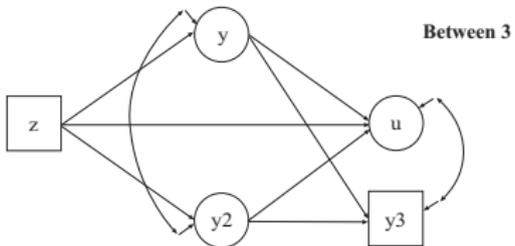
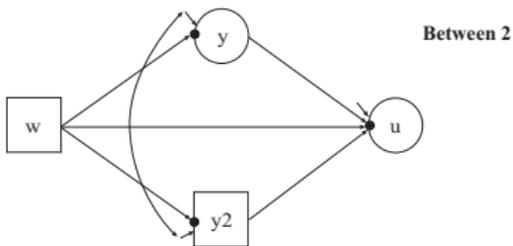
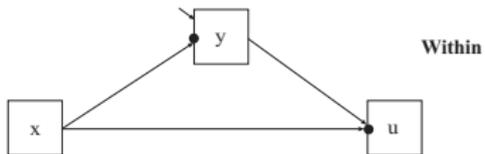
Model Results for Nurses Data, Continued

	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

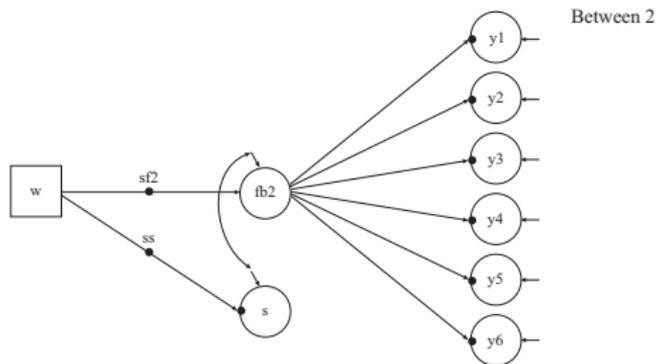
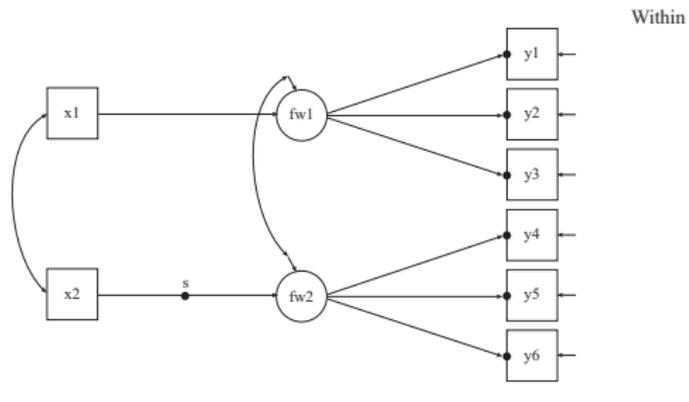
3-Level Path Analysis: UG Example 9.21



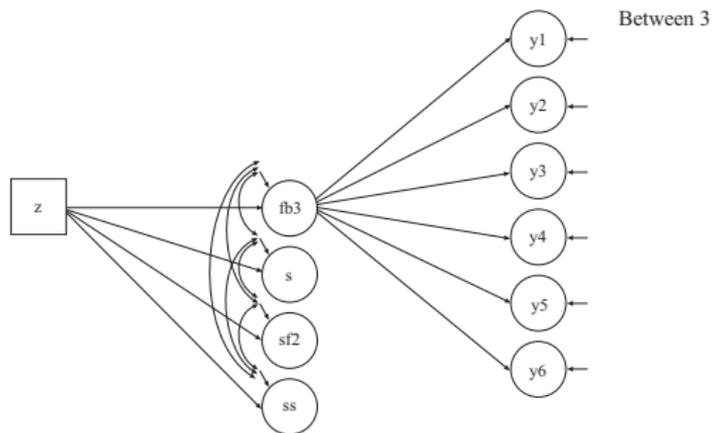
3-Level Path Analysis: UG Ex 9.21 Input

```
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;
```

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 = THREELLEVEL 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.

3-Level MIMIC Analysis, Monte Carlo Output, Continued

		ESTIMATES			S. E.	M. S. E.	95%	% Sig
Population		Average	Std. Dev.	Average		Cover	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

3-Level MIMIC Analysis, Monte Carlo Output, Continued

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

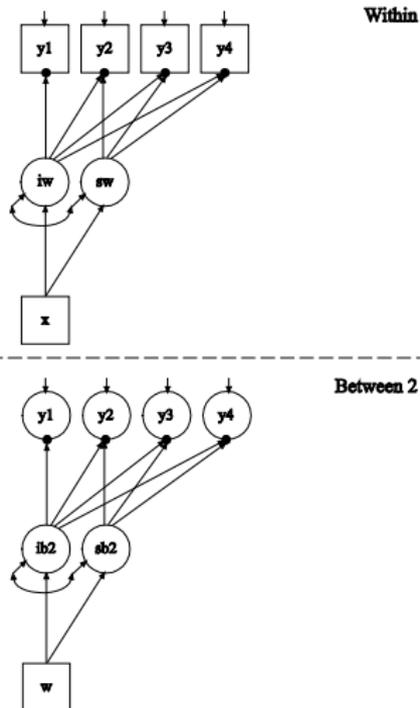
Intercepts

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

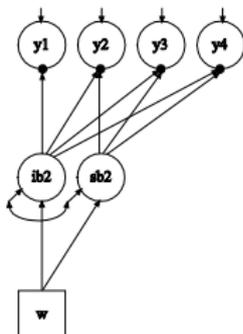
Residual Variances

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

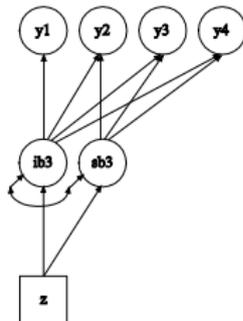
3-Level Growth Analysis



3-Level Growth Analysis, Continued



Between 2



Between 3

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 = THREELLEVEL;
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;
```

- 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

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

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

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

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

- 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 = threellevel; 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;
```

- 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)

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.

- 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

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,

$$y_{ijk} = \beta_0 + \beta_1 x_{ijk} + \beta_{2a j} + \beta_{2b k} + \varepsilon_{ijk}, \quad (8)$$

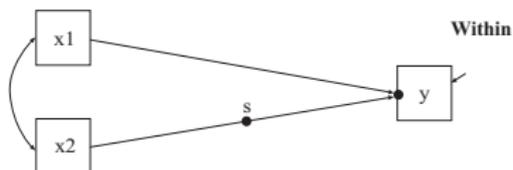
$$\beta_{2a j} = \gamma_{2a} w_{2a j} + \zeta_{2a j}, \quad (9)$$

$$\beta_{2b k} = \gamma_{2b} z_{2b k} + \zeta_{2b k}, \quad (10)$$

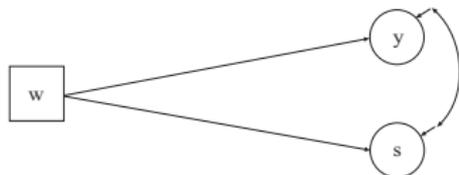
where

- 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

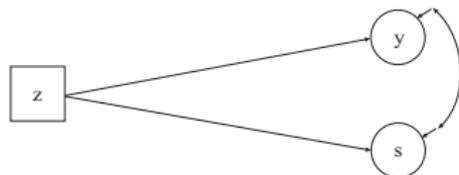
Cross-Classified Regression: UG Example 9.24



Level 2a



Level 2b



Cross-Classified Regression: Input for UG Example 9.24

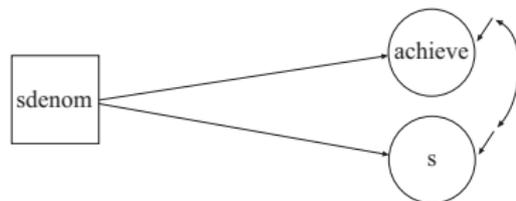
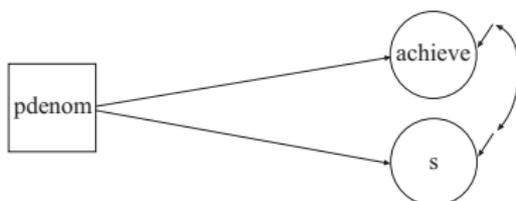
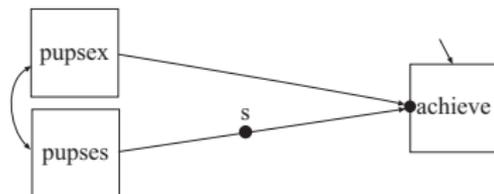
```
TITLE:      this is an example of a two-level
            regression for a continuous dependent
            variable using cross-classified data
DATA:      FILE = ex9.24.dat;
VARIABLE:  NAMES = y x1 x2 w z level2a level2b;
            CLUSTER = level2b level2a;
            WITHIN = x1 x2;
            BETWEEN = (level2a) w (level2b) z;
ANALYSIS:  TYPE = CROSSCLASSIFIED RANDOM;
            ESTIMATOR = BAYES;
            PROCESSORS = 2;
            ITERATIONS = (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;
```

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

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I.		Significance
				Lower 2.5%	Upper 2.5%	
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 puses
pdenom sdenom;
USEVARIABLES = achieve pupsex puses;
CLUSTER = pschool sschool;
WITHIN = pupsex puses;

ANALYSIS: ESTIMATOR = BAYES;
TYPE = CROSSCLASSIFIED;
PROCESSORS = 2;
FBITER = 5000;

MODEL: %WITHIN%
achieve ON pupsex puses;
%BETWEEN pschool%
achieve;
%BETWEEN school%
achieve;

OUTPUT: TECH1 TECH8;

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I.		Significance
				Lower 2.5%	Upper 2.5%	
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 puses
pdenom sdenom;
USEVARIABLES = achieve pupsex puses;
CLUSTER = pschool sschool;
WITHIN = pupsex puses;

ANALYSIS: ESTIMATOR = BAYES;
TYPE = CROSSCLASSIFIED RANDOM ;
PROCESSORS = 2; FBITER = 5000;

MODEL: %WITHIN%
achieve ON pupsex;
s | achieve ON puses;
%BETWEEN PSCHOOL%
achieve;
s;
%BETWEEN SSCHOOL%
achieve;
s@0;

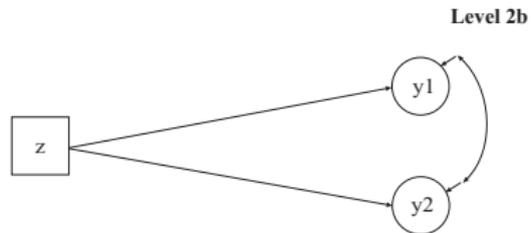
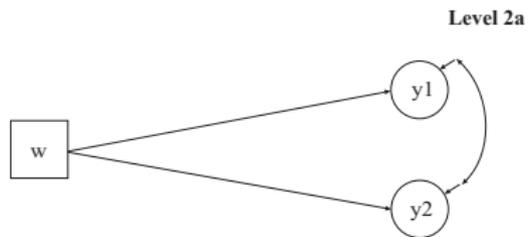
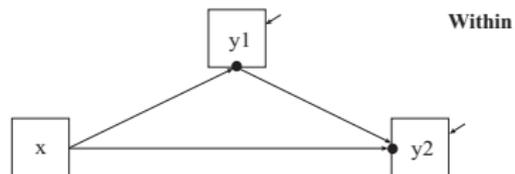
OUTPUT: TECH1 TECH8;

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I.		Significance
				Lower 2.5%	Upper 2.5%	
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.		Significance
				Lower 2.5%	Upper 2.5%	
WITHIN level						
achieve ON						
pupsex	0.261	0.047	0.000	0.168	0.351	*
pupes	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	*

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;
```