Latent Variable Modeling Using Mplus: Day 1

Bengt Muthén & Linda Muthén

Mplus
www.statmodel.com

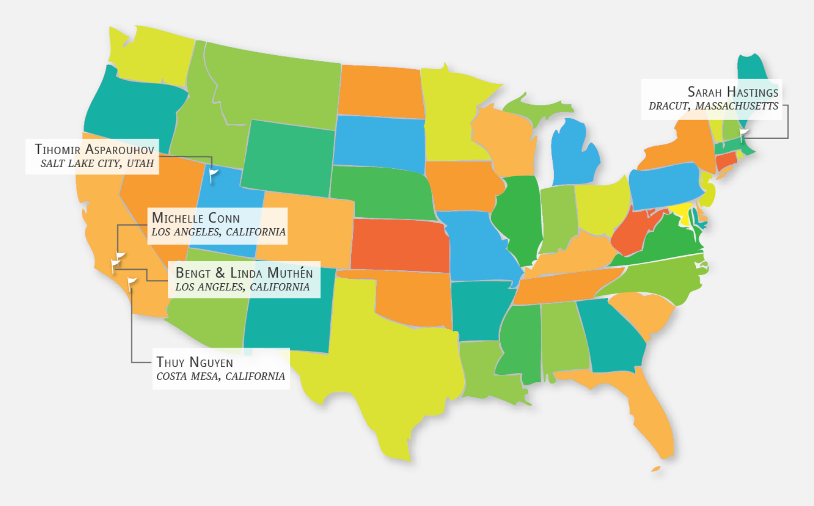
October, 2012

Table of Contents I

Table of Contents II

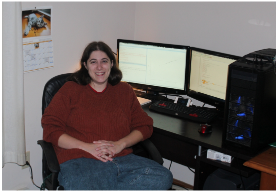
Table of Contents III

The Map Of The Mplus Team



Thuy

Michelle



Sarah

1. Mplus Background

- Inefficient dissemination of statistical methods:
  - Many good methods contributions from biostatistics, psychometrics, etc are underutilized in practice
- Fragmented presentation of methods:
  - Technical descriptions in many different journals
  - Many different pieces of limited software
- Mplus: Integration of methods in one framework
  - Easy to use: Simple, non-technical language, graphics
  - Powerful: General modeling capabilities

Mplus Integrates A Multitude Of Analysis Types Using The Unifying Theme Of Latent Variables

- Exploratory factor analysis
- Structural equation modeling
- Item response theory analysis
- Growth modeling
- Latent class analysis
- Latent transition analysis (Hidden Markov modeling)
- Growth mixture modeling
- Survival analysis
- Missing data modeling
- Multilevel analysis
- Complex survey data analysis
- Bayesian analysis
- Causal inference

Mplus Integrates A Multitude Of Analysis Types Using The Unifying Theme Of Latent Variables

- Exploratory factor analysis
- Structural equation modeling

Mplus Integrates A Multitude Of Analysis Types Using The Unifying Theme Of Latent Variables

- Structural equation modeling

- Bayesian analysis

Mplus Integrates A Multitude Of Analysis Types Using The Unifying Theme Of Latent Variables

- Survival analysis

- Latent class analysis

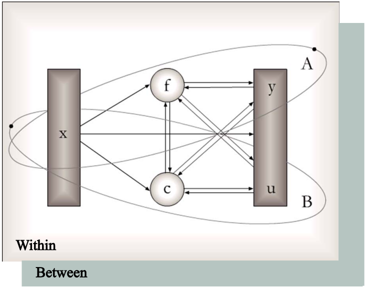
Mplus Integrates A Multitude Of Analysis Types Using The Unifying Theme Of Latent Variables

- Growth mixture modeling
- Survival analysis
- Missing data modeling

Mplus Integrates A Multitude Of Analysis Types Using The Unifying Theme Of Latent Variables

- Latent class analysis

- Causal inference



- Observed variables
  - x background variables (no model structure)
  - y continuous and censored outcome variables
  - u categorical (dichotomous, ordinal, nominal) and count outcome variables
- Latent variables
  - f continuous variables
    - interactions among fs
  - c categorical variables
    - multiple cs

Topics For Day 1 And Day 2 By Latent Variable Type

	Latent Variable Type	
Analysis	Continuous	Categorical
Path analysis		
Two-level path analysis	X	
Factor analysis	X	
Two-level factor analysis	X	
Structural equation modeling	X	
Growth modeling	X	
Count regression		X
Complier average causal effects		X
Latent class analysis		X
Factor mixture modeling	X	X
Latent transition analysis		X
Latent class growth analysis		X
Growth mixture modeling	X	X
Missing data modeling	X	X
Survival modeling	X	X

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

2. Mediation Path Analysis

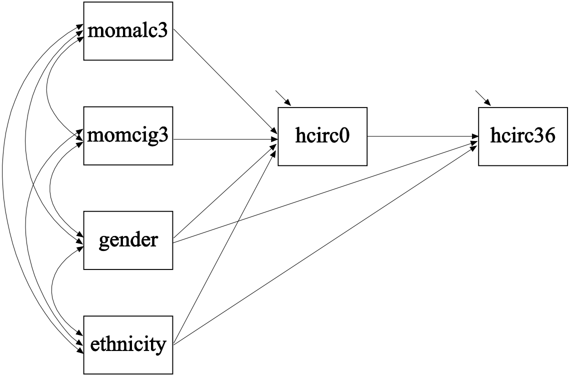
- 2.1 A simple mediation example: Fetal alcohol syndrome
- 2.2 Moderated mediation example: Aggressive classroom behavior
  - Version 7 LOOP plot of moderated mediation
- 2.3 Causally-defined effects in mediation analysis
- 2.4 Two-level path analysis with a binary outcome: High school dropout

The data are taken from the Maternal Health Project (Nancy Day). The subjects were a sample of mothers who drank at least three drinks a week during their first trimester plus a random sample of mothers who used alcohol less often.

Mothers were measured at the fourth and seventh month of pregnancy, at delivery, and at 8, 18, and 36 months postpartum. Offspring were measured at 0, 8, 18 and 36 months.

Data for the analysis include mothers' alcohol and cigarette use in the third trimester and the child's gender, ethnicity, and head circumference both at birth and at 36 months.

Fetal Alcohol Syndrome Example: Mediation Model



Input For Fetal Alcohol Syndrome Mediation Model

```
TITLE:      Fetal Alcohol Syndrome Mediation Model
DATA:       FILE = headalln.dat;
            FORMAT = 1f8.2 47f7.2;
VARIABLE:   NAMES = id weight0 weight8 weight18 weigh36 height0
            height8 height18 height36 hcirc0 hcirc8 hcirc18 hcirc36 mo-
            malc1 momalc2 momalc3 momalc8 momalc18 momalc36
            momcig1 momcig2 momcig3 momcig8 momcig18 momcig36
            gender eth momht gestage age8 age18 age36 esteem8 es-
            teem18 esteem36 faminc0 faminc8 faminc18 faminc36 mom-
            drg36 gravid sick8 sick18 sick36 advp advm1 advm2 advm3;
            MISSING = ALL (999);
            USEVARIABLES = momalc3 momcig3 hcirc0 hcirc36 gender
            eth;
            USEOBSERVATIONS = id NE 1121 AND NOT (momalc1 EQ
            999 AND momalc2 EQ 999 AND momalc3 EQ 999);
```

Input For Fetal Alcohol Syndrome Mediation Model, Continued

DEFINE:	hcirc0 = hcirc1/10;
	hcirc36 = hcirc36/10;
	momalc3 = log(momalc3 +1);
MODEL:	**hcirc36 ON hcirc0 gender eth;**
	hcirc0 ON momalc3 momcig3 gender eth;
MODEL INDIRECT:	hcirc36 IND hcirc0 momalc3;
	hcirc36 IND hcirc0 momcig3;
OUTPUT:	SAMPSTAT STANDARDIZED;

Output For Fetal Alcohol Syndrome Mediation Model

Test of Model Fit		
Chi-Square Test of Model Fit		
Value	1.781	
Degrees of Freedom	2	
P-Value	.4068	
RMSEA (Root Mean Square Error Of Approximation)		
Estimate	.000	
90 Percent C.I.	.000	0.079
Probability RMSEA <= .05	.774	

Output For Fetal Alcohol Syndrome Mediation Model, Continued

Model results

Parameter	Estimates	S.E.	Est./S.E.	Std	StdYX
hcirc36 ON					
hcirc0	.415	.036	11.382	.415	.439
gender	.762	.107	7.146	.762	.270
eth	-.094	.107	-.879	-.094	-.033
hcirc0 ON					
momalc3	-.500	.239	-2.090	-.500	-.084
momcig3	-.013	.005	-2.604	-.013	-.108
gender	.495	.118	4.185	.495	.166
eth	.578	.125	4.625	.578	.194

Output For Fetal Alcohol Syndrome Mediation Model, Continued

Model results

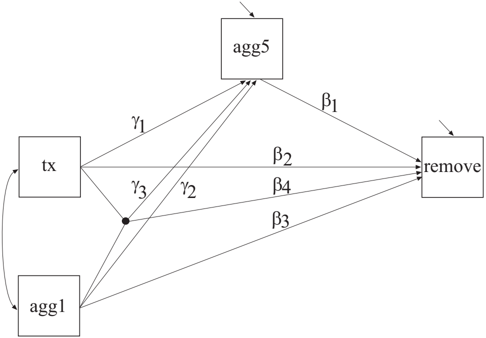
Parameter	Estimates	S.E.	Est./S.E.	Std	StdYX
Residual variances					
hcirc0	2.043	.119	17.107	2.043	.920
hcirc36	1.385	.087	15.844	1.385	.697
Intercepts					
hcirc0	33.729	.112	301.357	33.729	22.629
hcirc36	35.338	1.227	28.791	35.338	25.069

Standardized Indirect Effects

	Estimates	S.E.	Est./S.E.	Two-Tailed P-Value
Effects from MOMALC3 to HCIRC36				
Sum of indirect	-0.037	0.018	-2.047	0.041
Specific indirect				
HCIRC36				
HCIRC0				
MOMALC3	-0.037	0.018	-2.047	0.041
Effects from MOMCIG3 to HCIRC36				
Sum of indirect	-0.047	0.019	-2.557	0.011
Specific indirect				
HCIRC36				
HCIRC0				
MOMCIG3	-0.047	0.019	-2.557	0.011

2.2 Example: Moderated Mediation Of Aggressive Behavior

- Randomized field experiment in Baltimore public schools
- Classroom-based intervention in Grade 1 aimed at reducing aggressive-disruptive classroom behavior among elementary school students
- The variable agg1 represents the pre-intervention aggression score in Grade 1 used as a covariate in the analysis to strengthen the power to detect treatment effects
- Agg 1 also serves to explore a hypothesis of treatment-baseline interaction using the interaction between the treatment dummy variable tx and agg1, labeled inter. The agg1 covariate is referred to as a moderator
- The mediator variable agg5 is the Grade 5 aggression score
- The distal outcome variable remove is the number of times the student has been removed from school
- The analysis is based on $n = 392$ boys in treatment and control classrooms



$$remove = \beta_0 + \beta_1\, agg5 + \beta_2\, tx + \beta_3\, agg1 + \beta_4\, agg1\, tx + \varepsilon_1, \quad (1)$$

$$agg5 = \gamma_0 + \gamma_1\, tx + \gamma_2\, agg1 + \gamma_3\, agg1\, tx + \varepsilon_2, \quad (2)$$

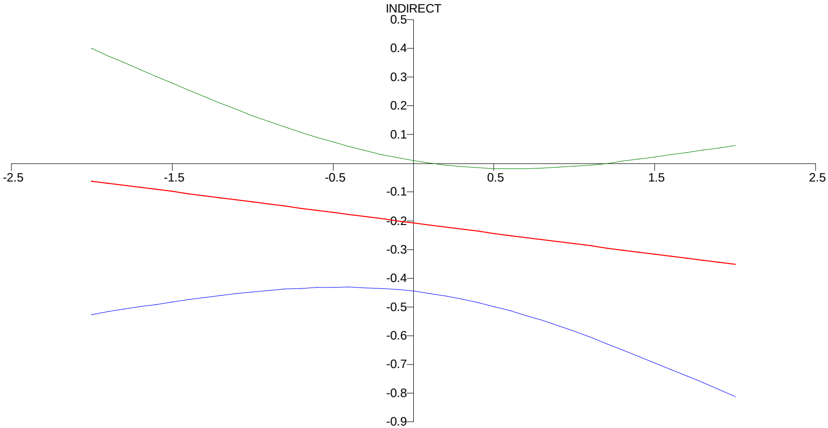
$$= \gamma_0 + (\gamma_1 + \gamma_3\, agg1)\, tx + \gamma_2\, agg1 + \varepsilon_2. \quad (3)$$

Indirect effect of tx on remove is $\beta_1\, (\gamma_1 + \gamma_3\, agg1)$, where agg1 moderates the effect of the treatment. Direct effect: $\beta_2 + \beta_4\, agg1$.

Input For Moderated Mediation Of Aggressive Behavior

```
DEFINE:          inter = tx*agg1;
ANALYSIS:        ESTIMATOR = BAYES;
                 PROCESSORS = 2; BITERATIONS = (30000);
MODEL:           remove ON agg5 (beta1)
                 tx (beta2)
                 agg1 (beta3)
                 inter (beta4);
                 agg5 ON tx (gamma1)
                 agg1 (gamma2)
                 inter (gamma3);
MODEL CONSTRAINT:
                 PLOT(indirect direct);
                 ! let moderate represent the range of the agg1 moderator
                 LOOP(moderate, -2, 2, 0.1);
                 indirect = beta1*(gamma1+gamma3*moderate);
                 direct = beta2+beta4*moderate;
PLOT:            TYPE = PLOT2;
```

Indirect Effect Of Treatment As A Function Of SD Units Of The Moderator agg1



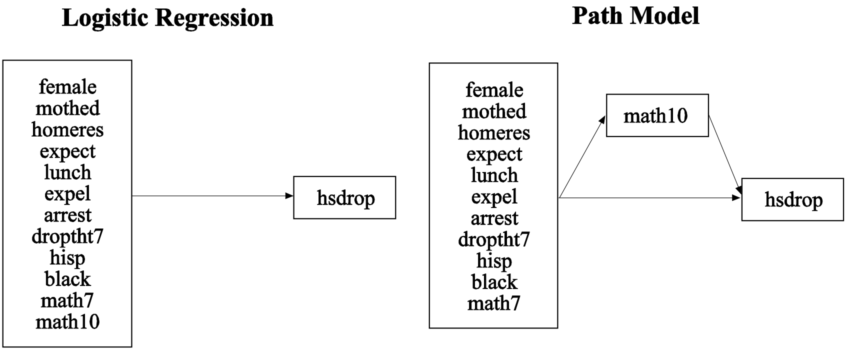
2.3 Causally-Defined Effects In Mediation Analysis

- Large, new literature on causal effect estimation: Robins, Greenland, Pearl, Holland, Sobel, VanderWeele, Imai
- New ways to estimate mediation effects with categorical and other non-normal mediators and distal outcomes
- Muthén (2011). Applications of Causally Defined Direct and Indirect Effects in Mediation Analysis using SEM in Mplus
  - The paper, an appendix with formulas, and Mplus scripts are available at www.statmodel.com under Papers, Mediational Modeling

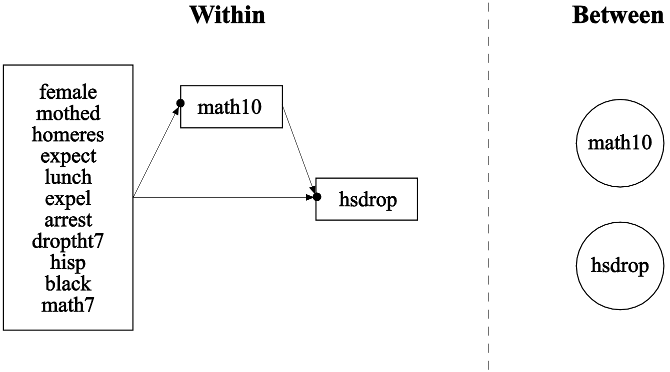
2.4 Two-Level Path Analysis With A Binary Outcome: High School Dropout

- Longitudinal Study of American Youth
- Math and science testing in grades 7 - 12
- Interest in high school dropout
- Data for 2,213 students in 44 public schools

A Path Model With A Binary Outcome And A Mediator With Missing Data



Two-Level Path Analysis With Random Intercepts



Input For A Two-Level Path Analysis Model With Random Intercepts, A Categorical Outcome, And Missing Data On The Mediating Variable

TITLE:	a twolevel path analysis with a categorical outcome and missing data on the mediating variable
DATA:	FILE = lsayfull_dropout.dat;
VARIABLE:	NAMES = female mothed homeres math7 math10 expel arrest hisp black hsdrop expect lunch droptht7 schcode;
	CATEGORICAL = hsdrop;
	CLUSTER = schcode;
	WITHIN = female mothed homeres expect math7 lunch expel arrest droptht7 hisp black;
ANALYSIS:	TYPE = TWOLEVEL;
	ESTIMATOR = ML;
	ALGORITHM = INTEGRATION;
	INTEGRATION = MONTECARLO (500);

Input For A Two-Level Path Analysis Model With Random Intercepts, A Categorical Outcome, And Missing Data On The Mediating Variable (Continued)

```
MODEL:     %WITHIN%
           hsdrop ON female mothed homeres expect math7 math10 lunch expel
           arrest droptht7 hisp black;
           math10 ON female mothed homeres expect math7 lunch expel arrest
           droptht7 hisp black;
           %BETWEEN%
           hsdrop math10;
OUTPUT:    PATTERNS SAMPSTAT STANDARDIZED TECH1 TECH8;
```

Output For A Two-Level Path Analysis Model With A Categorical Outcome And Missing Data On The Mediating Variable

Number of patterns	2
Number of clusters	44

Size (s)	Cluster ID with Size s	
12	304	
13	305	
36	307	122
38	106	112
39	138	109
40	103	
41	308	
42	146	120
43	102	101
44	303	143

Output For A Two-Level Path Analysis Model With A Categorical Outcome And Missing Data On The Mediating Variable (Continued)

Size (s)	Cluster ID with Size s					
45	141					
46	144					
47	140					
49	108					
50	126	111	110			
51	127	124				
52	137	117	147	118	301	136
53	142	131				
55	145	123				

Output For A Two-Level Path Analysis Model With A Categorical Outcome And Missing Data On The Mediating Variable (Continued)

Size (s)	Cluster ID with Size s	
57	135	105
58	121	
59	119	
73	104	
89	302	
93	309	
118	115	

Output For A Two-Level Path Analysis Model With A Categorical Outcome And Missing Data On The Mediating Variable (Continued)

Parameter	Estimate	S.E.	Est./S.E	Std	StdYX
Within Level					
hsdrop ON					
female	0.323	0.171	1.887	0.323	0.077
mothed	-0.253	0.103	-2.457	-0.253	-0.121
homeres	-0.077	0.055	-1.401	-0.077	-0.061
expect	-0.244	0.065	-3.756	-0.244	-0.159
math7	-0.011	0.015	-0.754	-0.011	-0.055
math10	-0.031	0.011	-2.706	-0.031	-0.197
lunch	0.008	0.006	1.324	0.008	0.074
expel	0.947	0.225	4.201	0.947	0.121
arrest	0.068	0.321	0.212	0.068	0.007
droptht7	0.757	0.284	2.665	0.757	0.074
hisp	-0.118	0.274	-0.431	-0.118	-0.016
black	-0.086	0.253	-0.340	-0.086	-0.013

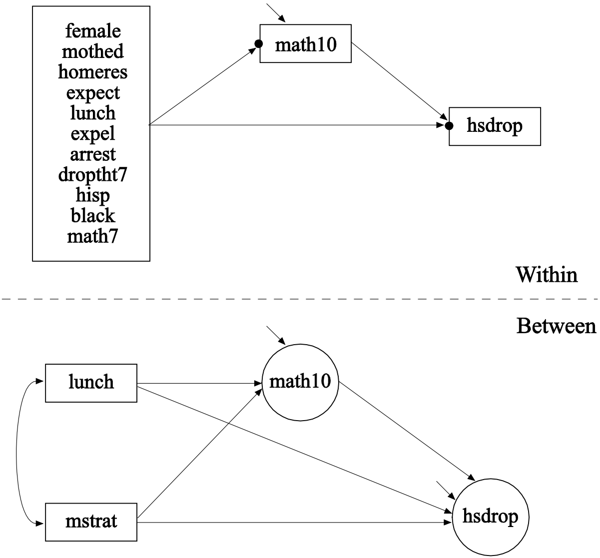
Output For A Two-Level Path Analysis Model With A Categorical Outcome And Missing Data On The Mediating Variable (Continued)

Parameter	Estimate	S.E.	Est./S.E	Std	StdYX
math10 ON					
female	-0.841	0.398	-2.110	-0.841	-0.031
mothed	0.263	0.215	1.222	0.263	0.020
homeres	0.568	0.136	4.169	0.568	0.070
expect	0.985	0.162	6.091	0.985	0.100
math7	0.940	0.023	40.123	0.940	0.697
lunch	-0.039	0.017	-2.308	-0.039	-0.059
expel	-1.293	0.825	-1.567	-1.293	-0.026
arrest	-3.426	1.022	-3.353	-3.426	-0.054
droptht7	-1.424	1.049	-1.358	-1.424	-0.022
hisp	-0.501	0.728	-0.689	-0.501	-0.010
black	-0.369	0.733	-0.503	-0.369	-0.009

Output For A Two-Level Path Analysis Model With A Categorical Outcome And Missing Data On The Mediating Variable (Continued)

Parameter	Estimate	S.E.	Est./S.E	Std	StdYX
Residual variances					
math10	62.010	2.162	28.683	62.010	0.341
Between Level					
Means					
math10	10.226	1.340	7.632	10.226	5.276
Thresholds					
hsdrop$1	-1.076	0.560	-1.920		
Variances					
hsdrop	0.286	0.133	2.150	0.286	1.000
math10	3.757	1.248	3.011	3.757	1.000

Two-Level Path Analysis Model Variation



3. Bayesian Analysis

- Bayesian analysis firmly established and its use is growing in mainstream statistics
- Much less use of Bayes outside statistics
- Bayesian analysis not sufficiently accessible in other programs
- Bayesian analysis was introduced in Mplus Version 6 and greatly expanded in Version 7: Easy to use
- Bayes provides a broad platform for further Mplus development

Why do we have to learn about Bayes?

- More can be learned about parameter estimates and model fit
- Better small-sample performance, large-sample theory not needed
- Priors can better reflect substantive hypotheses
- Analyses can be made less computationally demanding
  - Frequentists can see Bayes with non-informative priors as a computing algorithm to get answers that would be the same as ML if ML could have been done
- New types of models can be analyzed

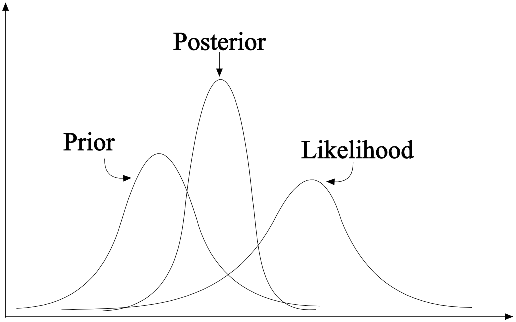
Writings On The Bayes Implementation In Mplus

- Asparouhov & Muthén (2010). Bayesian analysis using Mplus: Technical implementation. Technical Report. Version 3.
- Asparouhov & Muthén (2010). Bayesian analysis of latent variable models using Mplus. Technical Report. Version 4.
- Asparouhov & Muthén (2010). Multiple imputation with Mplus. Technical Report. Version 2.
- Asparouhov & Muthén (2010). Plausible values for latent variable using Mplus. Technical Report.
- Muthén (2010). Bayesian analysis in Mplus: A brief introduction. Technical Report. Version 3.
- Muthén & Asparouhov (2012). Bayesian SEM: A more flexible representation of substantive theory. Psychological Methods
- Asparouhov & Muthén (2011). Using Bayesian priors for more flexible latent class analysis.

Posted under Papers, Bayesian Analysis and Latent Class Analysis

Prior, Likelihood, And Posterior

- Frequentist view: Parameters are fixed. ML estimates have an asymptotically-normal distribution
- Bayesian view: Parameters are variables that have a prior distribution. Estimates have a possibly non-normal posterior distribution. Does not depend on large-sample theory
  - Non-informative (diffuse) priors vs informative priors



Bayesian Estimation Obtained Iteratively Using Markov Chain Monte Carlo (MCMC) Algorithms

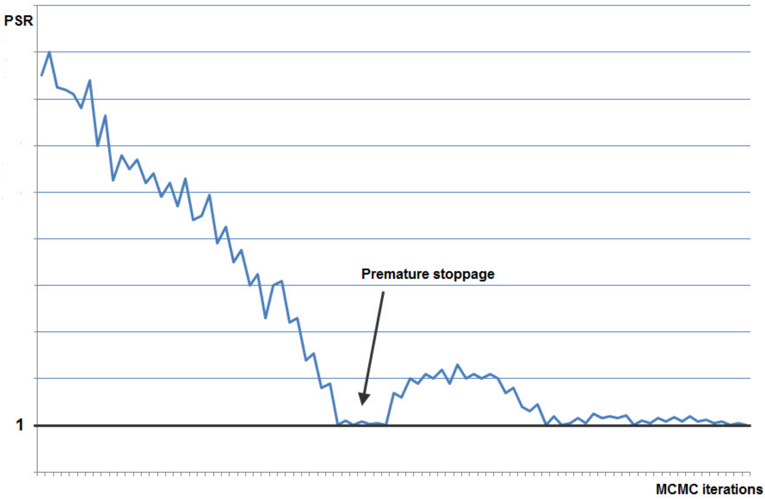
- θ_i: vector of parameters, latent variables, and missing observations at iteration i
- θ_i is divided into S sets:
  $\theta_i = (\theta_{1i}, ..., \theta_{Si})$
- Updated θ using Gibbs sampling over i = 1, 2, ..., n iterations:
  $\theta_{1i}|\theta_{2i-1}, ..., \theta_{Si-1}$, data, priors
  $\theta_{2i}|\theta_{3i-1}, ..., \theta_{Si-1}$, data, priors
  ...
  $\theta_{Si}|\theta_{1i}, ..., \theta_{S-1i-1}$, data, priors

Asparouhov & Muthén (2010). Bayesian analysis using Mplus. Technical implementation.Technical Report.

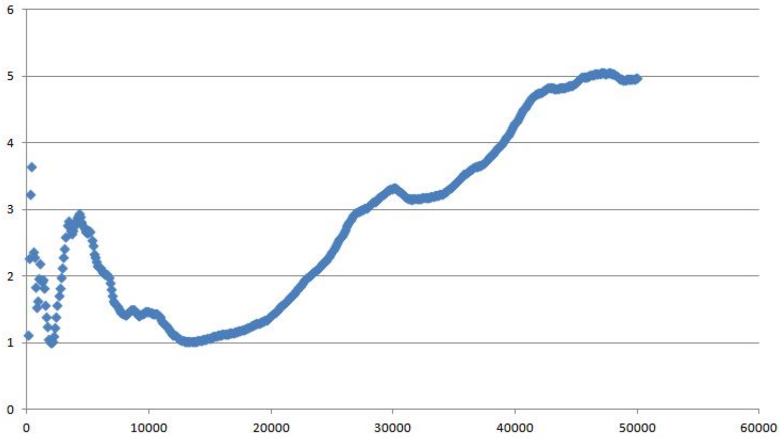
MCMC Iteration Issues

- Trace plot: Graph of the value of a parameter at different iterations
- Burnin phase: Discarding early iterations. Mplus discards first half
- Posterior distribution: Mplus uses the last half as a sample representing the posterior distribution
- Autocorrelation plot: Correlation between consecutive iterations for a parameter. Low correlation desired
- Mixing: The MCMC chain should visit the full range of parameter values, i.e. sample from all areas of the posterior density
- Convergence: Stationary process
- Potential Scale Reduction (PSR): Between-chain variation small relative to total variation. Convergence when PSR ≈ 1

PSR Convergence Issues: Premature Stoppage

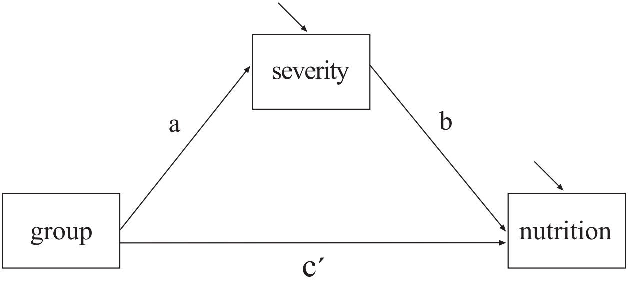


PSR Convergence Issues: Premature Stoppages Due to Non-Identification



3.1 Bayesian Mediation Modeling With Non-Informative Priors: The MacKinnon ATLAS Example

Source: MacKinnon et al. (2004), Multivariate Behavioral Research.
n = 861



- Intervention aimed at increasing perceived severity of using steroids among athletes. Perceived severity of using steroids is in turn hypothesized to increase good nutrition behaviors
- Indirect effect: $a \times b$

Input For Bayesian Analysis Of ATLAS Example Using The Default Of Non-Informative Priors

```
TITLE:        ATLAS
DATA:         FILE = mbr2004atlast.txt;
VARIABLE:     NAMES = obs group severity nutrit;
              USEVARIABLES = group - nutrit;
ANALYSIS:     ESTIMATOR = BAYES;
              PROCESSORS = 2;
              BITERATIONS = (10000); ! minimum of 10K iterations
MODEL:        severity ON group (a);
              nutrit ON severity (b)
              group;
MODEL CONSTRAINT:
              NEW (indirect);
              indirect = a*b;
OUTPUT:       TECH1 TECH8 STANDARDIZED;
PLOT:         TYPE = PLOT2;
```

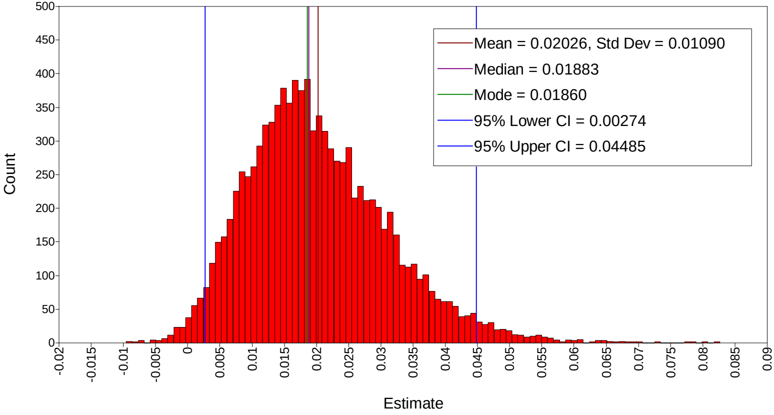
Output For Bayesian Analysis Of ATLAS Example

Parameter	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	95% C.I. Upper 2.5%
severity ON					
group	0.272	0.089	0.001	0.098	0.448
nutrit ON					
severity	0.074	0.030	0.008	0.014	0.133
group	-0.018	0.080	0.408	-0.177	0.140
Intercepts					
severity	5.648	0.062	0.000	5.525	5.768
nutrit	3.663	0.177	0.000	3.313	4.014

Output For Bayesian Analysis Of ATLAS Example (Continued)

Parameter	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	95% C.I. Upper 2.5%
Residual variances					
severity	1.719	0.083	0.000	1.566	1.895
group	1.333	0.065	0.000	1.215	1.467
New/additional parameters					
indirect	0.019	0.011	0.009	0.003	0.045

Bayesian Posterior Distribution For The Indirect Effect



Bayesian Posterior Distribution For The Indirect Effect: Conclusions

- Bayesian analysis: There is a mediated effect of the intervention
  - The 95% Bayesian credibility interval does not include zero
- ML analysis: There is not a mediated effect of the intervention
  - ML-estimated indirect effect is not significantly different from zero and the symmetric confidence interval includes zero
  - Bootstrap SEs and CIs can be used with ML

Types of factor analyses in Mplus:

- Exploratory Factor Analysis (EFA): Regular and bi-factor rotations
- Confirmatory Factor Analysis (CFA)
- Exploratory Structural Equation Modeling (ESEM; Asparouhov & Muthén, 2009 in Structural Equation Modeling)
- Bayesian Structural Equation Modeling (BSEM; Muthén & Asparouhov, 2012 in Psychological Methods)

Factor Analysis: Two Major Types

Factor analysis is a statistical method used to study the dimensionality of a set of variables. In factor analysis, latent variables represent unobserved constructs and are referred to as factors or dimensions.

- Exploratory Factor Analysis (EFA)
  Used to explore the dimensionality of a measurement instrument by finding the smallest number of interpretable factors needed to explain the correlations among a set of variables. Number of restrictions imposed: m^2, where m is the number of factors. Different rotations can be applied to find a simple factor loading pattern
- Confirmatory Factor Analysis (CFA)
  Used to study how well a factor model with hypothesized zero factor loadings fit the data. Number of restrictions imposed: $> m^2$. Rotation is avoided

Factor analysis is applied to a variety of measurement instruments:

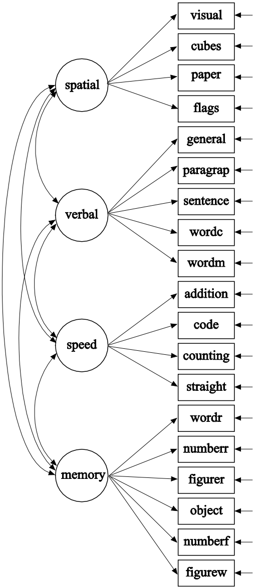
- Personality and cognition in psychology
  - Child Behavior Checklist (CBCL)
  - MMPI
- Attitudes in sociology, political science, etc.
- Achievement in education
- Diagnostic criteria in mental health

- Classic 1939 factor analysis study by Holzinger and Swineford (1939) in Illinois schools
- Twenty-six tests intended to measure a general factor and five specific factors
- Administered to seventh and eighth grade students in two schools
  - Grant-White school ($n = 145$). Students came from homes where the parents were mostly American-born
  - Pasteur school ($n = 156$). Students came largely from working-class parents of whom many were foreign-born and where their native language was used at home
- Source:
  - Holzinger, K. J. & Swineford, F. (1939). A study in factor analysis: The stability of a bi- factor solution. Supplementary Educational Monographs. Chicago, Ill.: The University of Chicago

Holzinger-Swineford Data, Continued

Current analyses:

- 19 variables using tests hypothesized to measure four mental abilities: Spatial, verbal, speed, and memory
- 24 variables, adding 5 tests measuring a general ability (deduction, test taking ability)



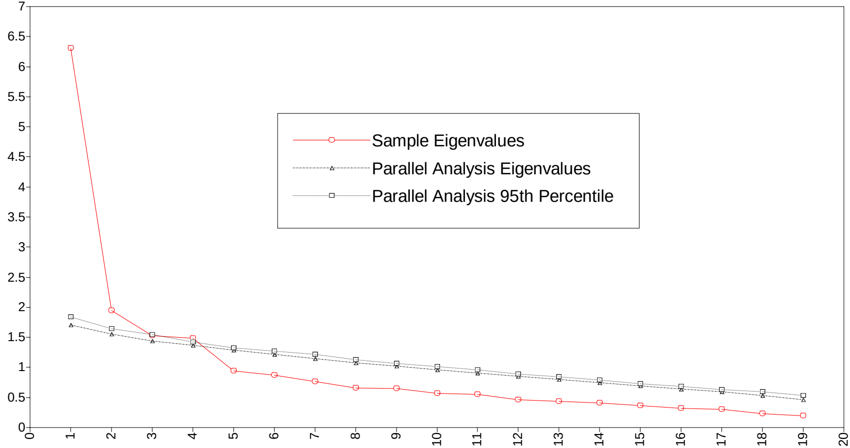
19 Variables: Expected Factor Loading Pattern

	Spatial	Verbal	Speed	Memory
visual	x	0	0	0
cubes	x	0	0	0
paper	x	0	0	0
flags	x	0	0	0
general	0	x	0	0
paragrap	0	x	0	0
sentence	0	x	0	0
wordc	0	x	0	0
wordm	0	x	0	0
addition	0	0	x	0
code	0	0	x	0
counting	0	0	x	0
straight	0	0	x	0
wordr	0	0	0	x
numberr	0	0	0	x
figurer	0	0	0	x
object	0	0	0	x
numberf	0	0	0	x
figurew	0	0	0	x

Holzinger-Swineford, 19 Variables: Input Excerpts For EFA

```
VARIABLE:    USEVARIABLES = visual - figurew;
             USEOBSERVATIONS = school EQ 0;
ANALYSIS:    TYPE = EFA 1 6;
             ROTATION = GEOMIN; ! default
             ESTIMATOR = ML; ! default
             PARALLEL = 50;
OUTPUT:      SAMPSTAT MODINDICES;
PLOT:        TYPE = PLOT3;
```

Parallel Analysis Of The Eigenvalues For 19-Variable Holzinger-Swineford, Grant-White EFA



EFA ML χ^2 Tests Of Model Fit For 19-Variable Holzinger-Swineford Data, Grant-White School

Factors	Chi-Square χ^2	df	p	BIC	CFI	RMSEA	SRMR
1	469.81	152	.000	18637	.68	.120	.102
2	276.44	134	.000	18534	.86	.086	.068
3	188.75	117	.000	18531	.93	.065	.053
4	110.34	101	.248	18532	.99	.025	.030
5	82.69	86	.581	18579	1.00	.000	.025
6	no convergence						

EFA ML Model Test Results For 4-Factor, 19-Variable Holzinger-Swineford Data For The Grant-White (n =145) And Pasteur (n=156) Schools

Model	χ^2	df	P-value	RMSEA	CFI
Grant-White					
EFA	110	101	0.248	0.025	0.991
Pasteur					
EFA	128	101	0.036	0.041	0.972

Estimated EFA factor pattern using oblique rotation with Geomin: Grant-White has 6 and Pasteur has 9 significant cross-loadings.

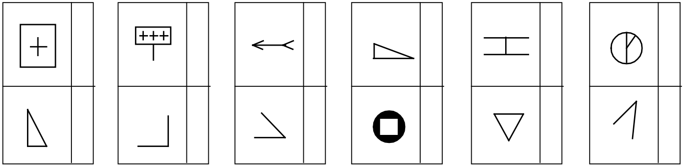
	Grant-White Factor Loading Patterns For EFA				Pasteur Factor Loading Pattern For EFA			
	Spatial	Verbal	Speed	Memory	Spatial	Verbal	Speed	Memory
visual	**0.628***	0.065	0.091	0.085	**0.580***	0.307*	-0.001	0.053
cubes	**0.485***	0.050	0.007	-0.003	**0.521***	0.027	-0.078	-0.059
paper	**0.406***	0.107	0.084	0.083	**0.484***	0.101	-0.016	-0.229*
flags	**0.579***	0.160	0.013	0.026	**0.687***	-0.051	0.067	0.101
general	0.042	**0.752***	0.126	-0.051	-0.043	**0.838***	0.042	-0.118
paragrap	0.021	**0.804***	-0.056	0.098	0.026	**0.800***	-0.006	0.069
sentence	-0.039	**0.844***	0.085	-0.057	-0.045	**0.911***	-0.054	-0.029
wordc	0.094	**0.556***	0.197*	0.019	0.098	**0.695***	0.008	0.083
wordm	0.004	**0.852***	-0.074	0.069	0.143*	**0.793***	0.029	-0.023
addition	-0.302*	0.029	**0.824***	0.078	-0.247*	0.067	**0.664***	0.026
code	0.012	0.050	**0.479***	0.279*	0.004	0.262*	**0.552***	0.082
counting	0.045	-0.159	**0.826***	-0.014	0.073	-0.034	**0.656***	-0.166
straight	0.346*	0.043	**0.570***	-0.055	0.266*	-0.034	**0.526***	-0.056
wordr	-0.024	0.117	-0.020	**0.523***	-0.005	0.020	-0.039	**0.726***
numberr	0.069	0.021	-0.026	**0.515***	-0.026	-0.057	-0.057	**0.604***
figurer	0.354*	-0.033	-0.077	**0.515***	0.329*	0.042	0.168	**0.403***
object	-0.195	0.045	0.154	**0.685***	-0.123	-0.005	0.333*	**0.469***
numberf	0.225	-0.127	0.246*	**0.450***	-0.014	0.092	0.092	**0.427***
figurew	0.069	0.099	0.058	**0.365***	0.139	0.013	0.237*	**0.291***

Factor Correlations For Grant-White And Pasteur Schools Using Oblique Geomin Rotation

	Grant-White Factor Correlations				Pasteur Factor Correlations			
	Spatial	Verbal	Speed	Memory	Spatial	Verbal	Speed	Memory
Spatial	1.000				1.000			
Verbal	0.378*	1.000			0.186*	1.000		
Speed	0.372*	0.386*	1.000		0.214*	0.326*	1.000	
Memory	0.307*	0.380*	0.375*	1.000	0.190*	0.100*	0.242*	1.000

- The item figurer is intended to measure the Memory ability factor but has a significant cross-loading on the Spatial ability factor for both the Grant-White and Pasteur schools
- Requires remembering a set of figures:

  Put a check mark ($\surd$) in the space after each figure that was on the study sheet. Do not put a check after any figure that you have not studied.



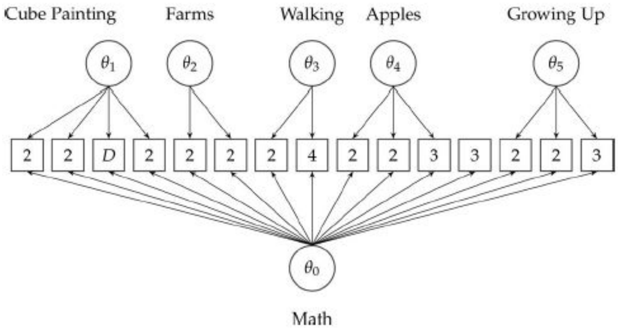
4.2 Bi-Factor Modeling Overview

- General factor influencing all items (deductive, test-taking ability); Holzinger-Swineford (1939) 24-variable model
- Testlet modeling, e.g. for PISA test items
- Longitudinal modeling with across-time correlation for residuals

Bi-factor modeling is as popular today as in 1939. New developments for faster maximum-likelihood estimation with categorical items, reducing the number of dimensions for numerical integration:

- Gibbons, & Hedeker (1992). Full-information item bi-factor analysis. Psychometrika
- Reise, Morizot, & Hays (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Quality of Life Research
- Cai (2010). A two-tier full-information item factor analysis model with applications. Psychometrika
- Cai, Yang, Hansen (2011). Generalized full-information item bifactor analysis. Psychological Methods

Bi-Factor Model For PISA Math Items



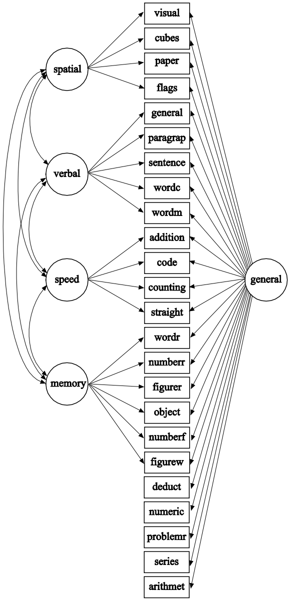
With categorical items, a two-tier algorithm for ML reduces the 6 dimensions of integration to 2.

Cai, Yang, & Hansen (2011) Generalized full-information item bifactor analysis. Psychological Methods, 16, 221-248

New Bi-Factor Modeling Methods

- Bi-factor EFA (Jennrich & Bentler, 2011, 2012, Psychometrika)
  - Allowing a general factor that influences all variables
  - ROTATION = BI-GEOMIN (new in Mplus Version 7)
- Bi-factor ESEM (Exploratory Structural Equation Modeling)
  - ROTATION = BI-GEOMIN (same as above)
  - Bi-factor ESEM with a general CFA factor and ROTATION = GEOMIN for specific factors
- Bi-factor BSEM (Bayesian SEM)
  - No rotation
  - Less rigid version of CFA bi-factor analysis

Holzinger-Swineford 24-variable bi-factor example:



	General	Spatial	Verbal	Speed	Memory
visual	x	x	0	0	0
cubes	x	x	0	0	0
paper	x	x	0	0	0
flags	x	x	0	0	0
general	x	0	x	0	0
paragrap	x	0	x	0	0
sentence	x	0	x	0	0
wordc	x	0	x	0	0
wordm	x	0	x	0	0
addition	x	0	0	x	0
code	x	0	0	x	0
counting	x	0	0	x	0
straight	x	0	0	x	0
wordr	x	0	0	0	x
numberr	x	0	0	0	x
figurer	x	0	0	0	x
object	x	0	0	0	x
numberf	x	0	0	0	x
figurew	x	0	0	0	x
deduct	x	0	0	0	0
numeric	x	0	0	0	0
problemr	x	0	0	0	0
series	x	0	0	0	0
arithmet	x	0	0	0	0

Bi-Factor Modeling Of The 24-Variable Holzinger-Swineford Data: Input Excerpts For Bi-Factor EFA

Requesting one general factor and four specific factors:

```
VARIABLE:    USEVARIABLES = visual - arithmet;
             USEOBSERVATIONS = school EQ 0;
ANALYSIS:    TYPE = EFA 5 5;
             ROTATION = BI-GEOMIN;
```

Bi-Factor EFA Solution For Holzinger-Swineford's 24-Variable Grant-White Data

	General	Spatial	Verbal	Speed	Memory
visual	**0.621***	**0.384***	-0.065	0.072	0.002
cubes	**0.433***	**0.207**	-0.103	-0.115	-0.118
paper	**0.430***	**0.343***	0.058	0.225	0.079
flags	**0.583***	**0.311***	-0.028	-0.077	-0.109
general	**0.610***	-0.034	**0.524***	0.001	-0.075
paragrap	**0.554***	0.053	**0.618***	0.012	0.102
sentence	**0.572***	-0.037	**0.622***	0.010	-0.064
wordc	**0.619***	0.006	**0.354***	0.038	-0.048
wordm	**0.582***	-0.008	**0.603***	-0.137	0.009
addition	**0.508***	-0.528	-0.036	**0.327**	0.009
code	**0.532***	-0.031	0.046	**0.428***	0.310*
counting	**0.568***	-0.229	-0.216*	**0.302**	-0.093
straight	**0.643***	0.217	0.004	**0.526***	-0.032

Bi-Factor EFA Results For Holzinger-Swineford, Continued

	General	Spatial	Verbal	Speed	Memory
wordr	**0.349***	0.018	0.077	0.032	**0.475***
numberr	**0.352***	0.037	-0.041	-0.052	**0.392***
figurer	**0.495***	0.221	-0.122	-0.033	**0.384***
object	**0.422***	-0.200	-0.010	-0.021	**0.497***
numberf	**0.553***	-0.041	-0.220*	0.003	**0.256***
figurew	**0.414***	-0.033	-0.003	-0.024	**0.246***
deduct	**0.611***	-0.001	0.089	-0.284*	0.036
numeric	**0.656***	-0.021	-0.129	0.029	-0.023
problemr	**0.607***	0.028	0.091	-0.227*	0.059
series	**0.714***	0.023	0.034	-0.202	-0.067
arithmet	**0.638***	-0.356*	0.092	-0.009	0.070

6 significant cross-loadings

BI-GEOMIN Factor Correlations

	General	Spatial	Verbal	Speed	Memory
General	1.000				
Spatial	0.000	1.000			
Verbal	0.000	0.022	1.000		
Speed	0.000	-0.223*	-0.122*	1.000	
Memory	0.000	-0.037	0.068	-0.134	1.000

ML χ^2 test of model fit has p-value = 0.3043.

Bi-Factor EFA Versus Regular EFA

- Bi-factor EFA with 1 general and m-1 specific factors has the same model fit as regular EFA with m factors (same ML loglikelihood and number of parameters); it is just another rotation of the factors
- For the 24-variable Holzinger-Swineford data, bi-factor EFA with 1 general and 4 specific factors gives a simple factor pattern that largely agrees with the Holzinger-Swineford hypotheses
- In contrast, regular 5-factor EFA for the 24-variable Holzinger-Swineford data does not give a simple factor loading pattern

4.3 The ESEM Factor Analysis Approach: Multiple-Group EFA Of Aggressive Behavior Of Males And Females

- 261 males and 248 females in Grade 3 (Baltimore Cohort 3)
- Teacher-rated aggressive-disruptive classroom behavior
- Outcomes treated as non-normal continuous variables
- Research question:
  - Does the measurement instrument function the same way for males and females?

Summary Of Separate Male/Female Exploratory Factor Analysis (Geomin Rotation)

	Loadings for Males			Loadings for Females		
Variables	Verbal	Person	Property	Verbal	Person	Property
Stubborn	**0.82***	-0.05	0.01	**0.88***	0.03	-0.22
Breaks Rules	**0.47***	**0.34***	0.01	**0.76***	0.06	-0.17
Harms Others and Property	-0.01	**0.63***	0.31*	**0.45***	0.03	0.36
Breaks Things	-0.02	0.02	**0.66***	-0.02	0.19	**0.43***
Yells At Others	**0.66***	0.23	-0.03	**0.97***	-0.23	0.05
Takes Other's Property	0.27*	0.08	**0.52***	0.02	**0.79***	0.10
Fights	0.22*	**0.75***	-0.00	**0.81***	-0.01	0.18
Harms Property	0.03	-0.02	**0.93***	0.27	0.20	**0.57***
Lies	**0.58***	0.01	0.27*	**0.42***	**0.50***	-0.00
Talks Back to Adults	**0.61***	-0.02	0.30*	**0.69***	0.09	-0.02
Teases Classmates	**0.46***	**0.44***	-0.04	**0.71***	-0.01	0.10
Fights With Classmates	0.30*	**0.64***	0.08	**0.83***	0.03	0.21*
Loses Temper	**0.64***	0.16*	0.04	**1.05***	-0.29	-0.01

Are The Factor Loading Patterns Significantly Different In The Different Groups?

Measurement invariance can be tested by multiple-group analysis

- But this involves a move from EFA to CFA
- CFA often premature
- CFA often rejected

- Why should we have to switch from EFA to CFA to test measurement invariance?

Staying With EFA: Multiple-Group Exploratory Factor Analysis (ESEM)

Asparouhov & Muthén (2009). Exploratory structural equation modeling. **Structural Equation Modeling**, 16, 397-438.

- Estimate by ML using a group-invariant unrotated factor loading matrix with a reference group having uncorrelated unit variance factors (m^2 restrictions), allowing group-varying factor covariance matrices and residual variances
- Rotate the common factor loading matrix, e.g. by oblique Geomin
- Transform the factor covariance matrices by the rotation matrix

- Factor loading invariance across groups can be tested by LR chi-square test: Not rejected for gender invariance

Multiple-Group EFA Modeling Results Using MLR

Model	LL0	C	# par. 's	Df	χ^2	CFI	RMSEA
M1	-8122	2.61	84	124	241	0.95	0.061
M2	-8087	2.41	94	114	188	0.97	0.050
M3	-8036	2.38	124	84	146	0.97	0.054

- M1: Loadings and intercepts invariance
- M2: Loadings but not intercepts invariance
- M3: Neither loadings nor intercepts invariance
- LL0: Log likelihood for the H0 (multiple-group EFA) model
- c is a non-normality scaling correction factor

Multiple-Group EFA Modeling Results Using MLR

- Comparing M3 and M2*:
  - cd = (94*2.41-124*2.38))/(-30) = 2.78
  - TRd = -2(LL0-LL1)/cd = 36.6 with 30 df: Loadings are invariant. Choose M2
- LL1 = loglikelihood for unrestricted H1 model (same for all 3) = -7934

* For loglikelihood difference testing with scaling corrections, see http://www.statmodel.com/chidiff.shtml

Male And Female Estimates From Multiple-Group EFA Using Invariant Factor Loadings (Standardized)

	Males			Females		
Variables	Verbal	Person	Property	Verbal	Person	Property
Stubborn	**0.80***	-0.01	-0.02	**0.86***	-0.00	-0.01
Breaks Rules	**0.53***	0.27*	0.01	**0.59***	0.20*	0.01
Harms Others & Property	0.00	**0.57***	0.35*	0.00	**0.56***	0.24*
Breaks Things	-0.01	-0.02	**0.67***	-0.03	-0.03	**0.63***
Yells At Others	**0.66***	0.25	-0.03	**0.69***	0.18	-0.01
Takes Others' Property	0.32*	0.04	**0.53***	**0.39***	0.03	**0.31***
Fights	0.28*	**0.74***	-0.03	0.35*	**0.61***	-0.02
Harms Property	0.11	0.03	**0.83***	0.19	0.04	**0.68***
Lies	**0.58***	0.01	0.30*	**0.67***	0.00	0.16*
Talks Back To Adults	**0.64***	-0.03	0.29*	**0.71***	-0.02	0.15*
Teases Classmates	**0.44***	**0.40***	0.02	**0.49***	**0.30***	0.01
Fights With Classmates	0.33*	**0.65***	0.05	0.41*	**0.53***	0.03
Loses Temper	**0.64***	0.19	0.00	**0.74***	0.14	0.00

Further ESEM Possibilities

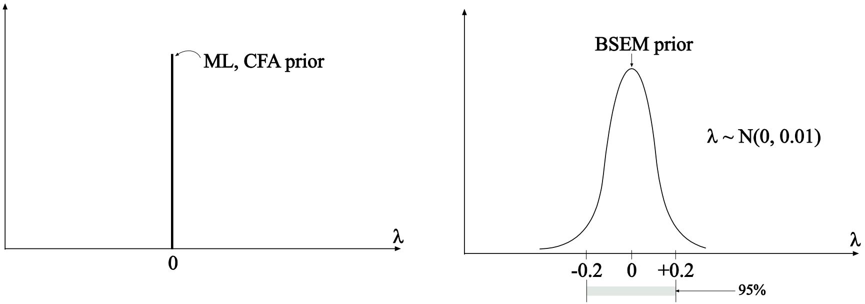
- Measurement intercept invariance testing and group differences in factor means
- Single-group invariance testing such as invariance across time with longitudinal factor analysis
- Exploratory SEM: EFA instead of or in combination with CFA measurement model

Asparouhov & Muthén (2009). Exploratory structural equation modeling. **Structural Equation Modeling**, 16, 397-438.

Muthén & Asparouhov (2010). Bayesian SEM: A more flexible representation of substantive theory. Psychological Methods, 17, 313-335.

- The BSEM paper
- 2 commentaries and a rejoinder
- Uses informative priors to estimate parameters that are not identified in ML

- ML CFA uses a very strong prior with an exact zero loading
- BSEM uses a zero-mean, small-variance prior for the loading:



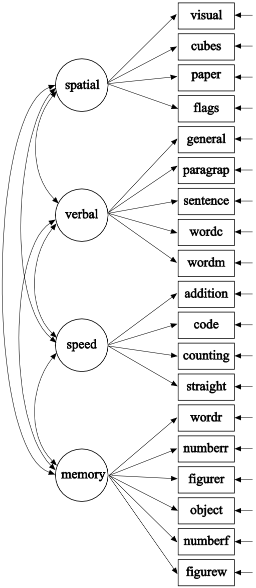
BSEM can be used to specify approximate zeros for

- Cross-loadings
- Residual correlations
- Direct effects from covariates
- Group and time differences in intercepts and loadings (new in Mplus Version 7)

Posterior Predictive Checking To Assess Model Fit And Sensitivity Analysis For Informative Priors

- Model fit based on a posterior predictive p-value (PPP; Gelman et al., 1996, Scheines et al., 1999) can be obtained via a fit statistic based on the usual chi-square test of H_0 against H_1. Low PPP indicates poor fit
- A 95% confidence interval is produced for the difference in chi-square for the real and replicated data; negative lower limit is good

- Sensitivity analysis is recommended for the choice of variance for the informative priors: How much do key parameters change as the prior variance is changed?
- As the variances of the informative priors are made larger, PPP increases and reaches a peak. SEs of estimates also increase and at some point the iterations won't converge (model is not identified)



CFA Factor Loading Pattern:

	Spatial	Verbal	Speed	Memory
visual	x	0	0	0
cubes	x	0	0	0
paper	x	0	0	0
flags	x	0	0	0
general	0	x	0	0
paragrap	0	x	0	0
sentence	0	x	0	0
wordc	0	x	0	0
wordm	0	x	0	0
addition	0	0	x	0
code	0	0	x	0
counting	0	0	x	0
straight	0	0	x	0
wordr	0	0	0	x
numberr	0	0	0	x
figurer	0	0	0	x
object	0	0	0	x
numberf	0	0	0	x
figurew	0	0	0	x

ML CFA Testing Results For Holzinger-Swineford Data For Grant-White (n =145) And Pasteur (n=156)

Model	χ^2	df	P-value	RMSEA	CFI
Grant-White					
CFA	216	146	0.000	0.057	0.930
EFA	110	101	0.248	0.025	0.991
Pasteur					
CFA	261	146	0.000	0.071	0.882
EFA	128	101	0.036	0.041	0.972

EFA has 6 (Grant-White) and 9 (Pasteur) significant cross-loadings

BSEM CFA For Holzinger-Swineford

- CFA: Cross-loadings fixed at zero - the model is rejected
- A more realistic hypothesis: Small cross-loadings allowed
- Cross-loadings are not all identified in terms of ML
- Different alternative: Bayesian CFA with informative priors for cross-loadings: $\lambda \sim$ N(0, 0.01).

  This means that 95% of the prior is in the range -0.2 to 0.2

Input Excerpts For BSEM CFA With 19 Items, 4 Factors, And Zero-Mean, Small-Variance Crossloading Priors

```
VARIABLE:   NAMES = id female grade agey agem school
            ! grade = 7/8
            ! school = 0/1 for Grant-White/Pasteur
            visual cubes paper flags general paragrap sentence wordc
            wordm addition code counting straight wordr numberr figurer
            object numberf figurew deduct numeric problemr series arith-
            met;
            USEV = visual-figurew;
            USEOBS = school eq 0;
DEFINE:     STANDARDIZE visual-figurew;
ANALYSIS:   ESTIMATOR = BAYES;
            PROCESSORS = 2;
            FBITER = 10000;
```

Input BSEM CFA 19 Items 4 Factors Crossloading Priors (Continued)

```
MODEL:        spatial BY visual* cubes paper flags;
              verbal BY general* paragrap sentence wordc wordm;
              speed BY addition* code counting straight;
              memory BY wordr* numberr figurer object numberf figurew;
              spatial-memory@1;
              ! cross-loadings:
              spatial BY general-figurew*0 (a1-a15);
              verbal BY visual-flags*0 (b1-b4);
              verbal BY addition-figurew*0 (b5-b14);
              speed BY visual-wordm*0 (c1-c9);
              speed BY wordr-figurew*0 (c10-c15);
              memory BY visual-straight*0 (d1-d13);
MODEL PRIORS:
              a1-d13 ~ N(0,0.01);
OUTPUT:       TECH1 TECH8 STDY;
PLOT:         TYPE = PLOT2;
```

ML analysis

Model	χ^2	Df	P-value	RMSEA	CFI
Grant-White					
CFA	216	146	0.000	0.057	0.930
EFA	110	101	0.248	0.025	0.991
Pasteur					
CFA	261	146	0.000	0.071	0.882
EFA	128	101	0.036	0.041	0.972

Bayesian analysis

Model	Sample LRT	2.5% PP limit	97.5% PP limit	PP p-value
Grant-White				
CFA	219	12	112	0.006
CFA w/ cross-loadings	142	-39	61	0.361
Pasteur				
CFA	264	56	156	0.000
CFA w/ cross-loadings	156	-28	76	0.162

	Grant-White Factor Loadings Using Informative Priors				Pasteur Factor Loadings Using Informative Priors			
	Spatial	Verbal	Speed	Memory	Spatial	Verbal	Speed	Memory
visual	**0.640***	0.012	0.050	0.047	**0.633***	0.145	0.027	0.039
cubes	**0.521***	-0.008	-0.010	-0.012	**0.504***	-0.027	-0.041	-0.030
paper	**0.456***	0.040	0.041	0.047	**0.515***	0.018	-0.024	-0.118
flags	**0.672***	0.046	-0.020	0.005	**0.677***	-0.095	0.026	0.093
general	0.037	**0.788***	0.049	-0.040	-0.056	**0.856***	0.027	-0.084
paragrap	-0.001	**0.837***	-0.053	0.030	0.015	**0.801***	-0.011	0.050
sentence	-0.045	**0.885***	0.021	-0.055	-0.063	**0.925***	-0.032	-0.036
wordc	0.053	**0.612***	0.096	0.029	0.055	**0.694***	0.013	0.063
wordm	-0.012	**0.886***	-0.086	0.020	0.092	**0.803***	0.001	0.012
addition	-0.172*	0.030	**0.795***	0.004	-0.147	-0.004	**0.655***	0.010
code	-0.002	0.054	**0.560***	0.130	-0.004	0.111	**0.655***	0.049
counting	0.013	-0.092	**0.828***	-0.049	0.025	-0.058	**0.616***	-0.057
straight	0.189*	0.043	**0.633***	-0.035	0.132	-0.067	**0.558***	0.001
wordr	-0.040	0.044	-0.031	**0.556***	-0.058	0.006	-0.090	**0.731***
numberr	0.003	-0.004	-0.038	**0.552***	0.006	-0.098	-0.106	**0.634***
figurer	0.132	-0.024	-0.049	**0.573***	0.156*	0.027	0.064	**0.517***
object	-0.139	0.014	0.029	**0.724***	-0.097	0.007	0.122	**0.545***
numberf	0.099	-0.071	0.095	**0.564***	-0.029	0.041	0.003	**0.474***
figurew	0.012	0.045	0.007	**0.445***	0.049	0.018	0.085	**0.397***

Number of significant cross-loadings: 2 for Grant-White and 1 for Pasteur

Sensitivity Analysis Using Different Variances For The Informative Priors Of The Cross-Loadings For The Holzinger-Swineford Data: Grant-White

Prior variance	95% cross-loading limit	PPP	Cross-loading (Posterior SD)	Factor corr. range
0.01	0.20	0.361	0.189 (.078)	0.443-0.557
0.02	0.28	0.441	0.248 (.096)	0.439-0.542
0.03	0.34	0.457	0.275 (.109)	0.432-0.530
0.04	0.39	0.455	0.292 (.120)	0.413-0.521
0.05	0.44	0.453	0.303 (.130)	0.404-0.513
0.06	0.48	0.447	0.309 (.139)	0.400-0.510
0.07	0.52	0.439	0.315 (.148)	0.395-0.508
0.08	0.55	0.439	0.319 (.156)	0.387-0.508
0.09	0.59	0.435	0.323 (.163)	0.378-0.506
0.10	0.62	0.427	0.327 (.171)	0.369-0.504

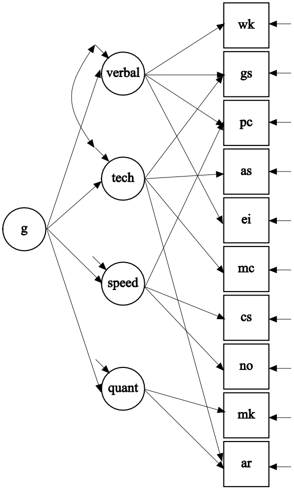
Summary Of Analyses Of Holzinger-Swineford 19-Variable Data

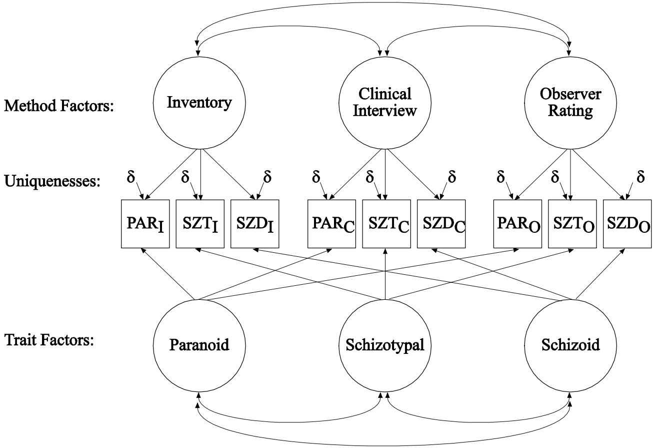
- Conventional, frequentist, CFA model rejected
- Bayesian CFA with informative cross-loadings not rejected
- The Bayesian approach uses an intermediate hypothesis:
  - Less strict than conventional CFA
  - Stricter than EFA, where the hypothesis only concerns the number of factors
  - Cross-loadings shrunken towards zero; acceptable degree of shrinkage monitored by PPP
- Bayes modification indices obtained by estimated cross-loadings
- Factor correlations: EFA $<$ BSEM $<$ CFA

Comparing BSEM And ESEM

- Similarities: Both ESEM and BSEM can be used for measurement models in SEM, including bi-factor models
- Differences:
  - ESEM is EFA-oriented while BSEM is CFA-oriented
  - ESEM uses a mechanical rotation and the rotation is not based on information from other parts of the model
  - BSEM is applicable not only to measurement models

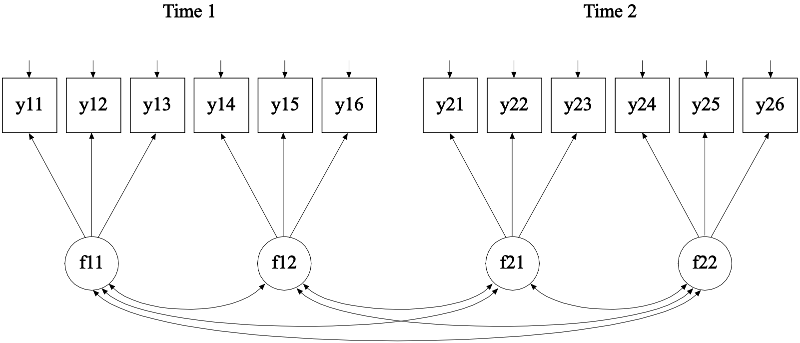
Model for the Armed Services Vocational Aptitude Battery (ASVAB):

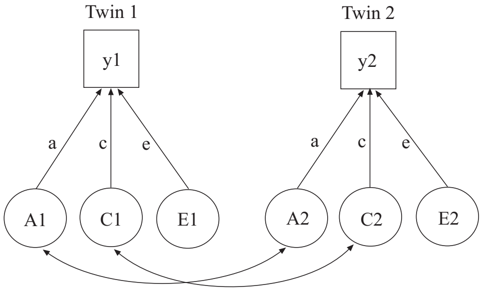




Source: Brown (2006)

4.5.3 Other Factor Models: Longitudinal Factor Analysis Model





- Continuous or categorical outcome
- MZ, DZ twins jointly in 2-group analysis

1.0 for MZ, 0.5 for DZ:

$$\Sigma_{DZ} = \begin{pmatrix} a^2 + c^2 + e^2 & \text{symm.} \\ 0.5 \times a^2 + c^2 & a^2 + c^2 + e^2 \end{pmatrix}$$

1.0:

$$\Sigma_{MZ} = \begin{pmatrix} a^2 + c^2 + e^2 & \text{symm.} \\ a^2 + c^2 & a^2 + c^2 + e^2 \end{pmatrix}$$

For Mplus inputs, see User's Guide ex5.18, ex5.21

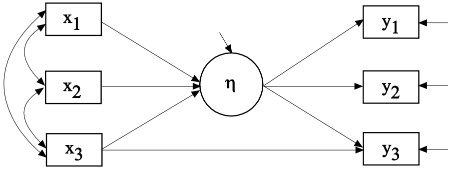
To further study a set of factors or latent variables established by a factor analysis, questions can be asked about the invariance of the measures and the heterogeneity of populations.

Measurement Invariance Does the factor model hold in other populations or at other time points?

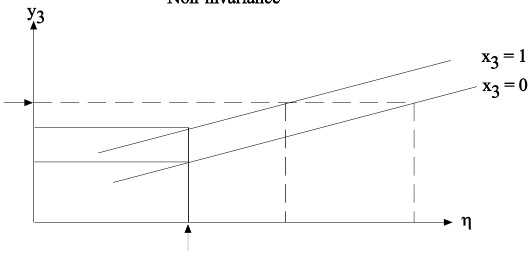
- Same number of factors
- Zero loadings in the same positions
- Equality of factor loadings
- Equality of intercepts
  - Test difficulty

Population Heterogeneity Are the factor means, variances, and covariances the same for different populations?

Non-invariance

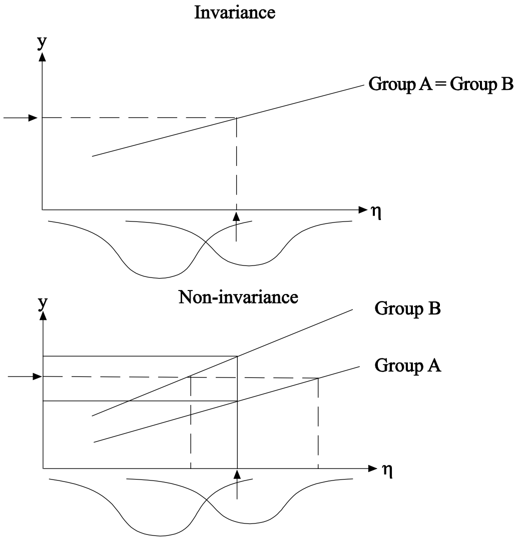


Non-invariance



Conditional on η, y is different for the two groups

Approach 2: Multiple-Group Analysis



Pros And Cons Of CFA With Covariates Versus Multiple-Group Analysis

- Advantages of CFA with covariates:
  - Easily handles many groups with small sample sizes
  - Parsimony: Only measurement intercepts represent non-invariance
  - Intercept non-invariance also for continuous (non-grouping) covariates
- Advantages of multiple-group analysis:
  - Allows factor loading non-invariance
  - Allows factor variance or item residual variance non-invariance

Multiple-group CFA with covariates possible.

5.1 CFA With Covariates (MIMIC): NELS Data

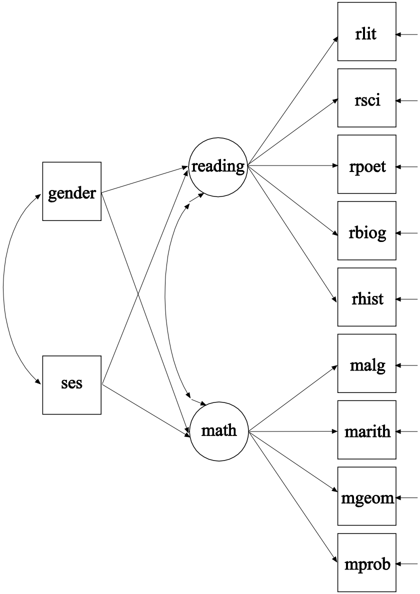
The NELS data consist of 16 testlets developed to measure the achievement areas of reading, math, science, and other school subjects. The sample consists of 4,154 eighth graders from urban, public schools.

Data for the analysis include five reading testlets and four math testlets. The entire sample is used.

Variables

rlit - reading literature
rsci - reading science
rpoet - reading poetry
rbiog - reading biography
rhist - reading history

malg - math algebra
marith - math arithmetic
mgeom - math geometry
mprob - math probability


rlit
rsci
rpoet
rbiog
rhist
malg
marith
mgeom
mprob
gender
ses
reading
math

Input For NELS CFA With Covariates

TITLE:	CFA with covariates using NELS data
DATA:	FILE = ft21.dat;
VARIABLE:	NAMES = ses rlit rsci rpoet rbiog rhist malg marith mgeom mprob searth schem slife smeth hgeog hcit hhist gender schoolid minorc;
	USEVARIABLES = rlit-mprob **ses gender**;
MODEL:	reading BY rlit-rhist;
	math BY malg-mprob;
	reading math ON ses gender; ! female = 0, male = 1
OUTPUT:	STANDARDIZED MODINDICES (3.84);

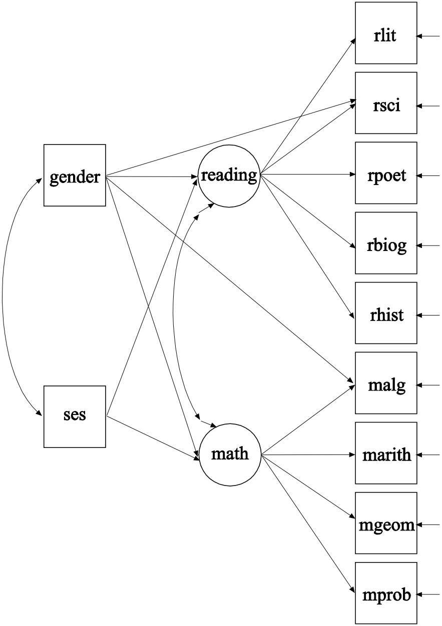
Output Excerpts For NELS CFA With Covariates

Model results

Parameter	Estimate	S.E.	Est./S.E.	Std	StdYX
reading ON					
ses	.344	.014	24.858	.407	.438
gender	-.186	.027	-6.901	-.220	-.110
math ON					
ses	.418	.015	28.790	.412	.444
gender	.044	.030	1.457	.044	022

Output Excerpts Modification Indices For Direct Effects NELS CFA With Covariates

	M.I.	E.P.C.	StdE.P.C.	StdYX E.P.C.
rsci ON gender	**31.730**	**0.253**	**0.253**	**0.073**
rpoet ON gender	12.715	-0.124	-0.124	-0.045
rhist ON ses	6.579	0.062	0.062	0.038
malg ON gender	**26.616**	**-0.120**	**-0.120**	**-0.051**
marith ON gender	10.083	0.075	0.075	0.032
mgeom ON ses	4.201	0.040	0.040	0.032
mprob ON gender	7.922	0.143	0.143	0.037


rlit
rsci
rpoet
rbiog
rhist
malg
marith
mgeom
mprob
gender
ses
reading
math

Output Excerpts NELS CFA With Covariates And Two Direct Effects

Parameter	Estimate	S.E.	Est./S.E.	Std	StdYX
reading ON					
ses	0.343	0.014	24.854	0.406	0.437
gender	-0.222	0.028	-7.983	-0.262	-0.131
math ON					
ses	0.419	0.015	28.807	0.411	0.444
gender	**0.092**	**0.032**	**2.873**	**0.090**	**0.045**
rsci ON					
gender	**0.254**	**0.045**	**5.649**	**0.254**	**0.073**
malg ON					
gender	**-0.121**	**0.023**	**-5.171**	**-0.121**	**-0.051**

Conclusions: Effects Related To The Math Factor

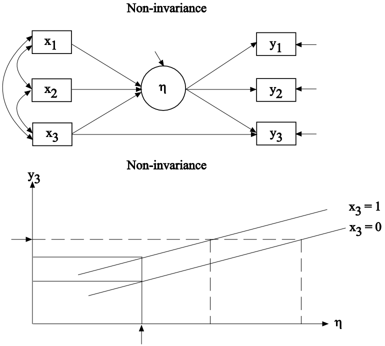
Gender effect on the math factor:

- Allowing no direct effects:
  - No significant gender effect
- Allowing direct effects:
  - Significant gender effect
- The positive gender effect on the math factor combined with the negative direct effect of gender on the malg item results in a non-significant gender effect on the math factor when ignoring measurement non-invariance

Partial measurement non-invariance is ok when modeled.

Interpretation Of The Positive Direct Effect Of Gender On rsci

- Direct effect is positive - for a given reading factor value, males do better than expected on rsci
- Conclusion - rsci is not invariant. Males may have had more exposure to science reading



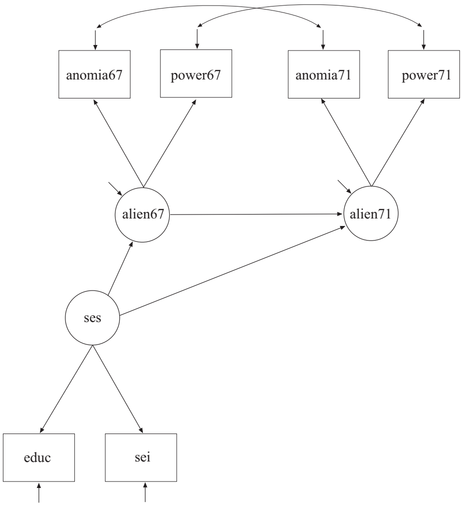
5.2 Multiple-Group Analysis

Mplus offers several alternative types of multiple-group analyses:

- Conventional multiple-group analysis based on measurement invariance for a CFA measurement model
- ESEM multiple-group analysis based on measurement invariance for an EFA measurement model
- BSEM multiple-group analysis based on a measurement model allowing approximate measurement invariance

These topics are not further discussed here. Day 3 touches on multiple-group examples.
Video and handouts covering multiple-group analysis are provided in Topic 1 as well as in the August 2012 Utrecht course; see the Mplus web site.

6. Structural Equation Modeling (SEM): Classic Wheaton Et Al. SEM



Input For Classic Wheaton Et Al. SEM

```
TITLE:     Classic structural equation model with multiple indicators used
           in a study of the stability of alienation.
DATA:      FILE = wheacov.dat;
           TYPE = COVARIANCE;
           NOBS = 932;
VARIABLE:  NAMES = anomia67 power67 anomia71 power71 educ sei;
MODEL:     ses BY educ sei;
           alien67 BY anomia67 power67;
           alien71 BY anomia71 power71;
           alien71 ON alien67 ses;
           alien67 ON ses;
           anomia67 WITH anomia71;
           power67 WITH power71;
OUTPUT:    SAMPSTAT STANDARDIZED MODINDICES (0);
```

Tests of model fit

Chi-Square Test of Model Fit		
Value	4.771	
Degrees of Freedom	4	
P-Value	.3111	
RMSEA (Root Mean Square Error of Approximation)		
Estimate	.014	
90 Percent C.I.	.000	.053
Probability RMSEA <= .05	.928	

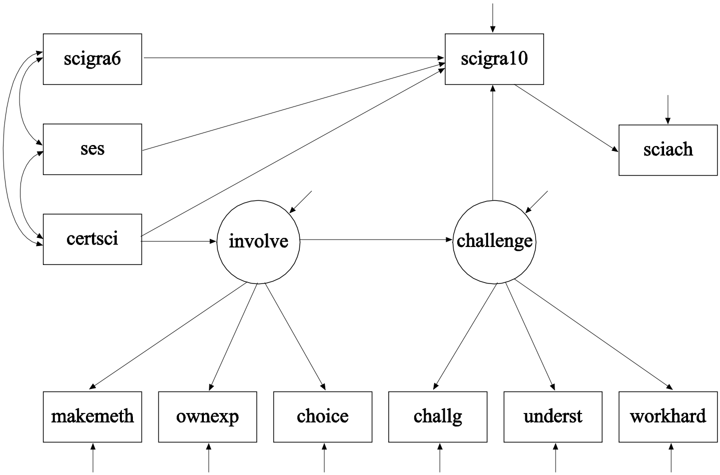
Model results

Parameter	Estimate	S.E.	Est./S.E.	Std	StdYX
ses BY					
educ	1.000	0.000	0.000	2.607	0.841
sei	5.221	0.422	12.367	13.612	0.642
alien67 BY					
anomia67	1.000	0.000	0.000	2.663	0.775
power67	0.979	0.062	15.896	2.606	0.852
alien71 BY					
anomia71	1.000	0.000	0.000	2.850	0.805
power71	0.922	0.059	15.500	2.627	0.832

Parameter	Estimate	S.E.	Est./S.E.	Std	StdYX
alien71 ON					
alien67	0.607	0.051	11.895	0.567	0.567
ses	-0.227	0.052	-4.337	-0.208	-0.208
alien67 ON					
ses	-0.575	0.056	-10.197	-0.563	-0.563
anomia67 WITH					
anomia71	1.622	0.314	5.173	1.622	0.356
power67 WITH					
power71	0.340	0.261	1..302	0.340	0.121

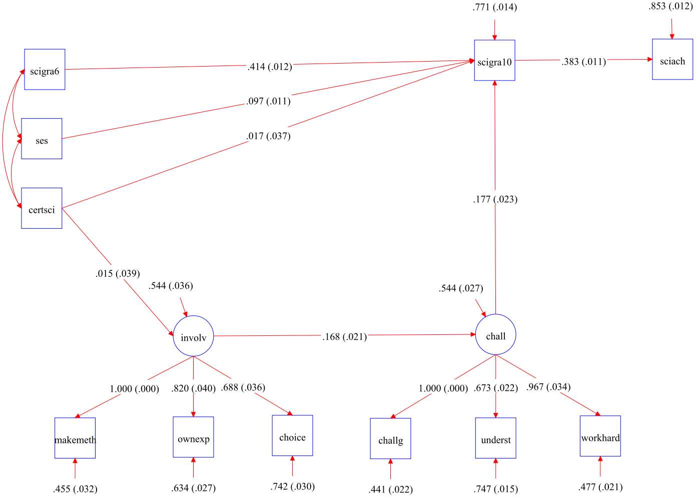
Output For Classic Wheaton Et Al. SEM (Continued)

Parameter	Estimate	S.E.	Est./S.E.	Std	StdYX
Residual variances					
anomia67	4.730	0.453	10.438	4.730	0.400
power67	2.564	0.403	6.362	2.564	0.274
anomia71	4.397	0.515	8.357	4..397	0.351
power71	3.072	0.434	7.077	3.072	0.308
educ	2.804	0.507	5.532	2.804	0.292
sei	264.532	18.125	14.595	264.532	0.588
alien67	4.842	0.467	10.359	0.683	0.683
alien71	4.084	0.404	10.104	0.503	0.503
Variances					
ses	6.796	0.649	10.476	1.000	1.000



Analyzed by BSEM in Muthén & Asparouhov (2012).

Kaplan Science SEM Using The Mplus Diagrammer



The Mplus Diagrammer

The Mplus Diagrammer can be used to draw

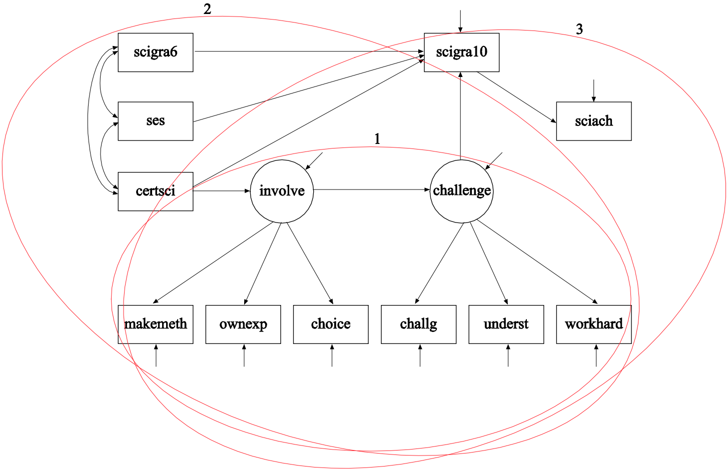
- An input diagram:
  Diagramming on the left, producing Mplus input on the right
- An output diagram
- A diagram using Mplus input without analysis:
  A new drawing tool

  Developed by Delian Asparouhov, Tihomir Asparouhov, and Thuy Nguyen

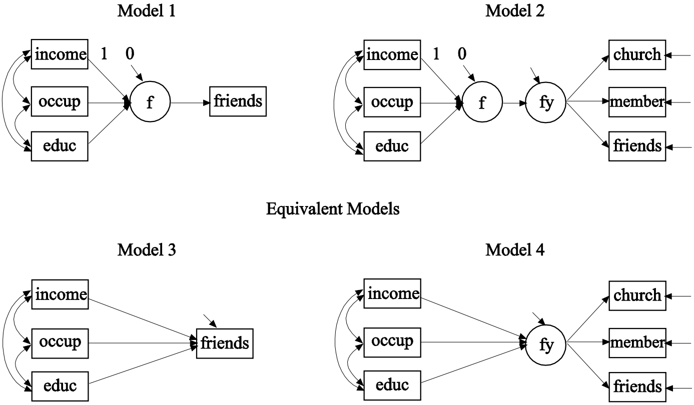
6.1 Modeling Issues In SEM

- Model building strategies
  - Bottom up
  - Measurement versus structural parts
- Number of indicators
  - Identifiability
  - Robustness to misspecification
- Believability
  - Measures
  - Direction of arrows
  - Other models
- Quality of estimates
  - Parameters, S.E.'s, power
  - Monte Carlo study within the substantive study

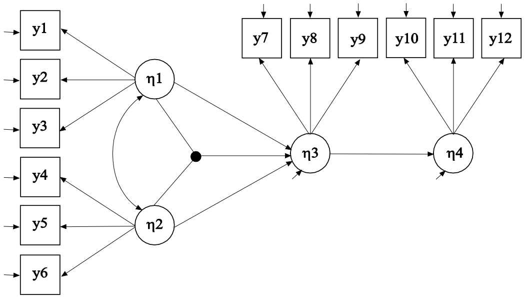
Kaplan Science SEM: Understanding The Parts Of The Model



Formative Indicators. Equivalent Models



Structural Equation Model With Interaction Between Latent Variables



- Mplus uses ML estimation and the XWITH option (Klein & Moosbrugger, 2000)
- Marsh et al. (2004) compares estimators
- FAQ at www.statmodel.com: Latent variable interactions discusses interpretation, variances, standardization, and plots

New SEM Features In Version 7

- 3-level analysis with a full SEM for each level (TYPE=THREELEVEL)
    - Continuous outcomes: ML and Bayes
    - Continuous and categorical outcomes: Bayes
- 4-level complex survey data (TYPE=COMPLEX THREELEVEL): Stratification, weights on all levels, 3 cluster variables)
- Cross-classified analysis with a full SEM for each level (TYPE=CROSSCLASSIFIED)
- 3-level and cross-classified multiple imputation

For other Version 7 news, see Version History at www.statmodel.com.

7. Growth Modeling: Typical Examples

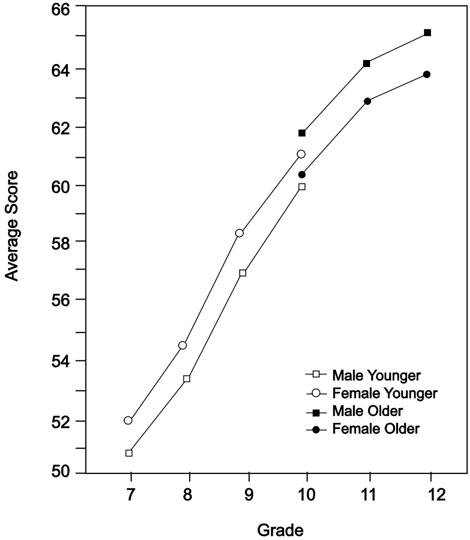
- Linear growth of achievement over grades: LSAY
- Non-linear growth of head circumference
- Multiple-indicator growth

LSAY Data

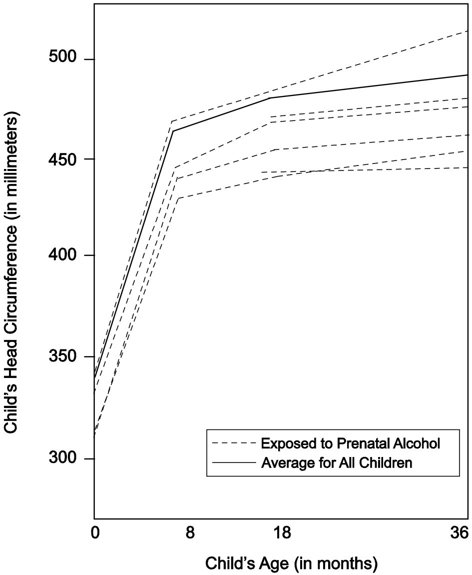
Longitudinal Study of American Youth (LSAY)

- Two cohorts measured each year beginning in 1987
  - Cohort 1 - Grades 10, 11, and 12
  - Cohort 2 - Grades 7, 8, 9, 10, 11, and 12
- Each cohort contains approximately 60 schools with approximately 60 students per school
- Variables - math and science achievement items, math and science attitude measures, and background variables from parents, teachers, and school principals
- Approximately 60 items per test with partial item overlap across grades - adaptive tests

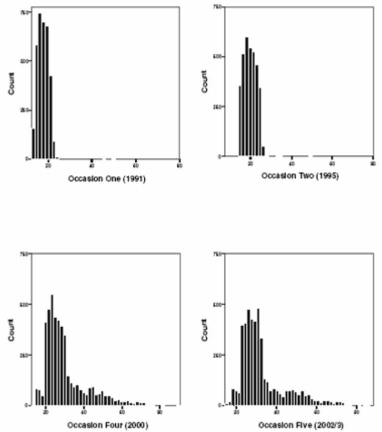
LSAY Math Total Score



Mothers' Alcohol Use And Offspring Head Circumference



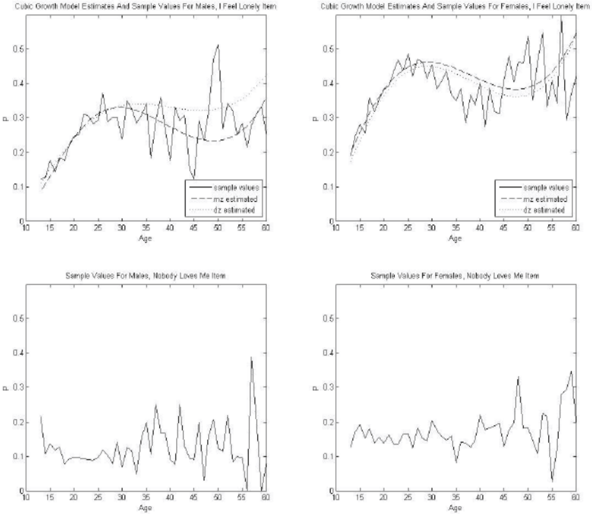
Loneliness In Twins



Age range: 13-85
5 occasions: 1991, 1995, 1997, 2000, 2002/3

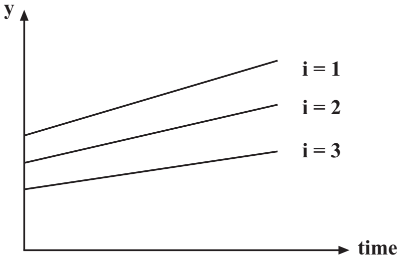
Boomsma, D.I., Cacioppo, J.T., Muthén, B., Asparouhov, T., & Clark, S. (2007). Longitudinal Genetic Analysis for Loneliness in Dutch Twins. Twin Research and Human Genetics, 10, 267-273.

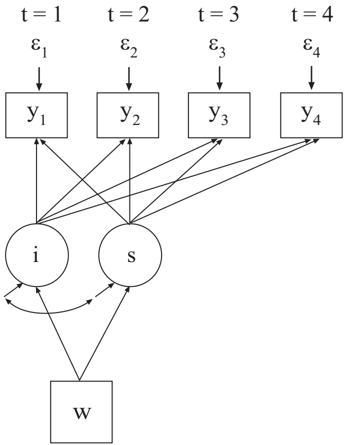
Males Females



I feel lonely

Nobody loves me





$$(1) \quad y_{ti} = i_i + s_i \; time_{ti} + \varepsilon_{ti}$$

$$(2a) \quad i_i = \alpha_0 + \gamma_0 \; w_i + \zeta_{0i}$$

$$(2b) \quad s_i = \alpha_1 + \gamma_1 \; w_i + \zeta_{1i}$$

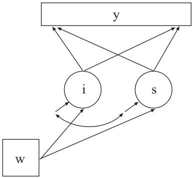
Growth Modeling Approached In Two Ways: Data Arranged As Wide Versus Long Format

- Wide: Multivariate, Single-Level Approach

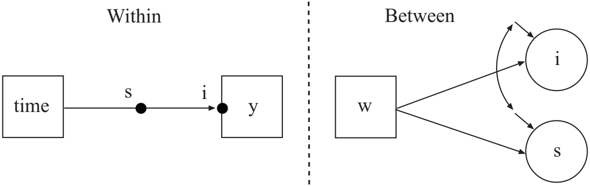
  $y_{ti} = i_i + s_i * \text{time}_{ti} + \varepsilon_{ti}$

  i_i regressed on w_i

  s_i regressed on w_i



- Long: Univariate, 2-Level Approach (CLUSTER = id)



The intercept i is called y in Mplus. See UG ex9.16.

Conventional Growth Modeling With Random Slopes: Long Format, Univariate, Two-Level

Time point t, individual i (two-level modeling, no clustering):

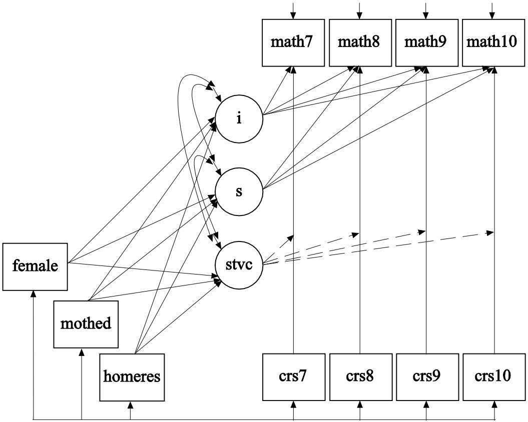
y_{ti}: repeated measures of the outcome, e.g. math achievement
a_{1ti}: time-related variable; e.g. grade 7-10
a_{2ti}: time-varying covariate, e.g. math course taking
x_i: time-invariant covariate, e.g. home background

Two-level analysis with random slopes for individually-varying times of observation and time-varying covariates:

$$\text{Level 1:} \quad y_{ti} = \pi_{0i} + \pi_{1i}\, a_{1ti} + \pi_{2ti}\, a_{2ti} + e_{ti}, \tag{4}$$

$$\text{Level 2:} \begin{cases} \pi_{0i} = \beta_{00} + \beta_{01}\, x_i + r_{0i}, \\ \pi_{1i} = \beta_{10} + \beta_{11}\, x_i + r_{1i}, \\ \pi_{2i} = \beta_{20} + \beta_{21}\, x_i + r_{2i}. \end{cases} \tag{5}$$

Growth Modeling With Random Slopes: Wide Format, Multivariate, Single-Level

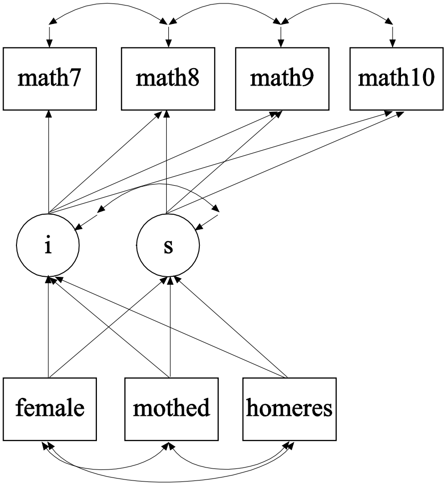


Pros And Cons Of Wide Versus Long

- Advantages of the wide approach:
  - Modeling flexibility
    - Unequal residual variances and covariances
    - Testing of measurement invariance with multiple indicator growth
    - Allowing partial measurement non-invariance
  - Missing data modeling
  - Reduction of the number of levels by one (or more)
- Advantages of the long approach
  - Can handle many time points

Advantages Of Growth Modeling In A Latent Variable Framework

- Flexible curve shape
- Individually-varying times of observation
- Regressions among random effects
- Multiple processes
- Modeling of zeroes
- Multiple populations
- Multiple indicators
- Embedded growth models
- Categorical latent variables: growth mixtures


math7
math8
math9
math10
i
s
female
mothed
homeres

Input Excerpts For LSAY Linear Growth Model With Time-Invariant Covariates

```
TITLE:      Growth 7 - 10, no covariates;
DATA:       FILE = lsayfull_dropout.dat;
VARIABLE:   NAMES = lsayid schcode female mothed homeres
            math7 math8 math9 math10 math11 math12
            mthcrs7 mthcrs8 mthcrs9 mthcrs10 mthcrs11 mthcrs12;
            MISSING = ALL (999);
            USEVAR = math7-math10 female mothed homeres;
ANALYSIS:   !ESTIMATOR = MLR;
MODEL:      i s | math7@0 math8@1 math9@2 math10@3;
            i s ON female mothed homeres;
Alternative language:
MODEL:      i BY math7-math10@1;
            s BY math7@0 math8@1 math9@2 math10@3;
            [math7-math10@0];
            [i s];
            i s ON female mothed homeres;
```

Output Excerpts For LSAY Linear Growth Model With Time-Invariant Covariates

n = 3116

Tests of model fit for ML

Chi-square test of model fit		
Value	33.611	
Degrees of freedom	8	
P-value	0.000	
CFI/TLI		
CFI	0.998	
TLI	0.994	
RMSEA (Root Mean Square Error of Approximation)		
Estimate	0.032	
90 Percent C.I.	0.021	0.044
Probability RMSEA <= .05	0.996	
SRMR (Standardized Root Mean Square Residual)		
Value	0.010	

Output Excerpts LSAY Growth Model With Time-Invariant Covariates (Continued)

Selected estimates for ML

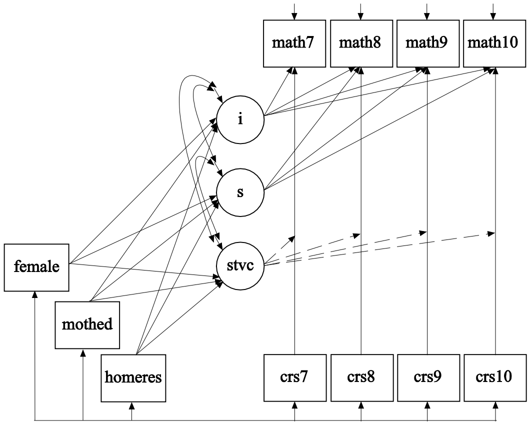
	Estimates	S.E.	Est./S.E.	Two-Tailed P-value
i ON				
female	2.123	0.327	6.489	0.000
mothed	2.262	0.164	13.763	0.000
homeres	1.751	0.104	16.918	0.000
s ON				
female	-0.134	0.116	-1.153	0.249
mothed	0.223	0.059	3.771	0.000
homeres	0.273	0.037	7.308	0.000

Output Excerpts LSAY Growth Model With Time-Invariant Covariates (Continued)

Selected estimates for ML

	Estimates	S.E.	Est./S.E.	Two-Tailed P-value
s WITH				
i	4.131	1.244	3.320	0.001
Residual variances				
i	71.888	3.630	19.804	0.000
s	3.313	0.724	4.579	0.000
Intercepts				
i	38.434	0.497	77.391	0.000
s	2.636	0.181	14.561	0.000

7.3 LSAY Growth Modeling With Random Slopes



Input For LSAY Growth Modeling With Random Slopes

TITLE:	Growth model with individually varying times of observation and random slopes
DATA:	FILE IS lsaynew.dat;
	FORMAT IS 3F8.0 F8.4 8F8.2 3F8.0;
VARIABLE:	NAMES ARE math7 math8 math9 math10 crs7 crs8 crs9 crs10 female mothed homeres a7-a10;
	! crs7-crs10 = highest math course taken during each
	! grade (0=no course, 1=low, basic, 2=average, 3=high.
	! 4=pre-algebra, 5=algebra I, 6=geometry,
	! 7=algebra II, 8=pre-calc, 9=calculus)

Input For LSAY Growth Modeling With Random Slopes (Continued)

	MISSING ARE ALL (9999);
	TSCORES = a7-a10;
DEFINE:	CENTER crs7-crs10 mothed homeres (GRANDMEAN);
	math7 = math7/10;
	math8 = math8/10;
	math9 = math9/10;
	math10 = math10/10;
ANALYSIS:	TYPE = **RANDOM** MISSING;
	ESTIMATOR = ML;
	MCONVERGENCE = .001;

Input For LSAY Growth Modeling With Random Slopes (Continued)

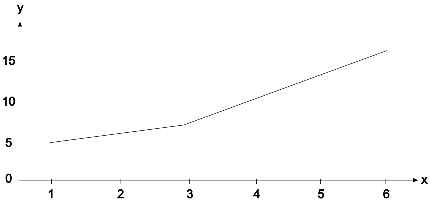
MODEL:	**i s | math7-math10 AT a7-a10;**
	stvc | math7 ON crs7;
	stvc | math8 ON crs8;
	stvc | math9 ON crs9;
	stvc | math10 ON crs10;
	i s stvc ON female mothed homeres;
	i WITH s;
	stvc WITH i;
	stvc WITH s;
OUTPUT:	TECH8;

7.4 Six Ways To Model Non-Linear Growth

- Estimated time scores
- Quadratic (cubic) growth model
- Fixed non-linear time scores
- Piecewise growth modeling
- Time-varying covariates
- Non-linearity of random effects*

* Grimm & Ram (2009). Nonlinear growth models in Mplus and SAS. Structural Equation Modeling, 16, 676-701.

- Can be used to represent different phases of development
- Can be used to capture non-linear growth
- Each piece has its own growth factor(s)
- Each piece can have its own coefficients for covariates



One intercept growth factor, two slope growth factors

s1:	0 1 2 2 2 2	Time scores piece 1
s2:	0 0 0 1 2 3	Time scores piece 2

Input Excerpts For Piecewise Growth Modeling

One intercept growth factor, two slope growth factors

s1: 0 1 2 2 2 2 Time scores piece 1

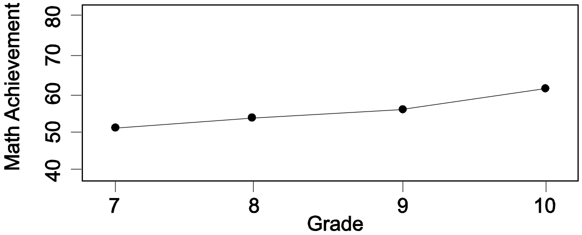
s2: 0 0 0 1 2 3 Time scores piece 2

```
VARIABLE:   USEVARIABLES = y1-y6;
MODEL:      i s1 | y1@0 y2@1 y3@2 y4@2 y5@2 y6@2;
            i s2 | y1@0 y2@0 y3@0 y4@1 y5@2 y6@3;
```

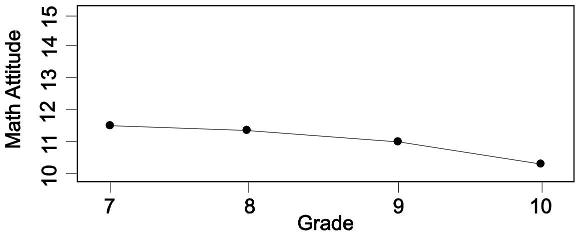
7.6 Growth Modeling With Multiple Processes

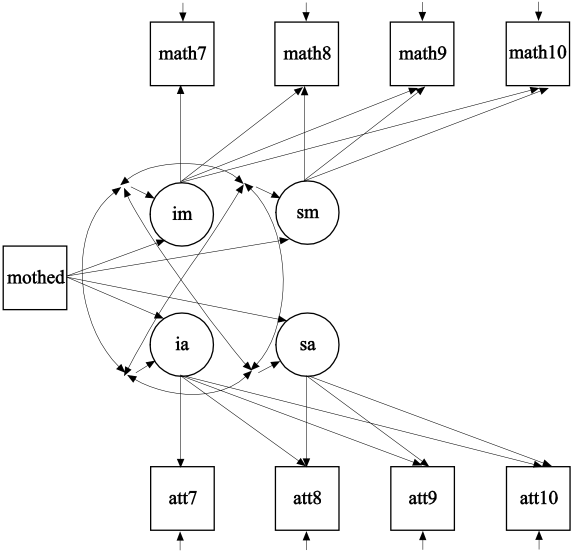
- Parallel processes
- Sequential processes

LSAY Sample Means for Math



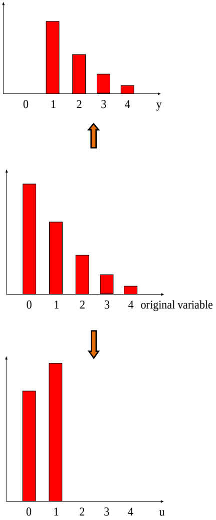
Sample Means for Attitude Towards Math

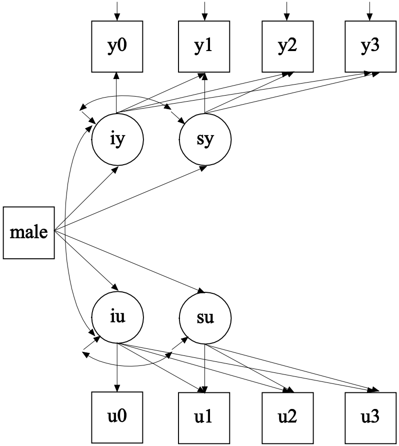



math7
math8
math9
math10
im
sm
mothed
ia
sa
att7
att8
att9
att10

Input For LSAY Parallel Process Growth Model

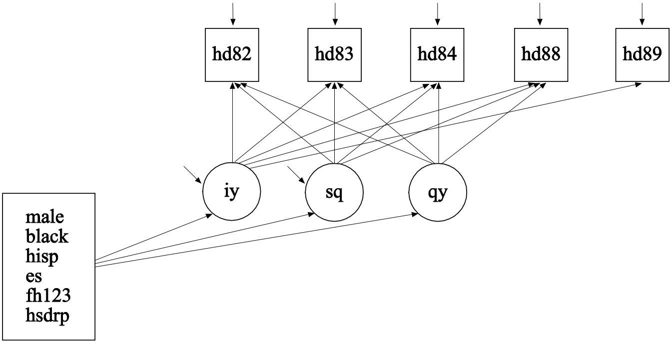
```
TITLE:      LSAY For Younger Females With Listwise Deletion
            Parallel Process Growth Model-Math Achievement and Math
            Attitudes
DATA:       FILE IS lsay.dat;
            FORMAT IS 3f8 f8.4 8f8.2 3f8 2f8.2;
VARIABLE:   NAMES ARE cohort id school weight math7 math8 math9
            math10 att7 att8 att9 att10 gender mothed homeres ses3 sesq3;
            USEOBS = (gender EQ 1 AND cohort EQ 2);
            MISSING = ALL (999);
            USEVAR = math7-math10 att7-att10 mothed;
MODEL:      im sm | math7@0 math8@1 math9 math10;
            ia sa | att7@0 att8@1 att9@2 att10@3;
            im-sa ON mothed;
```


0
1
2
3
4
y
0
1
2
3
4
original variable
0
1
2
3
4
u


y0
y1
y2
y3
iy
sy
male
iu
su
u0
u1
u2
u3

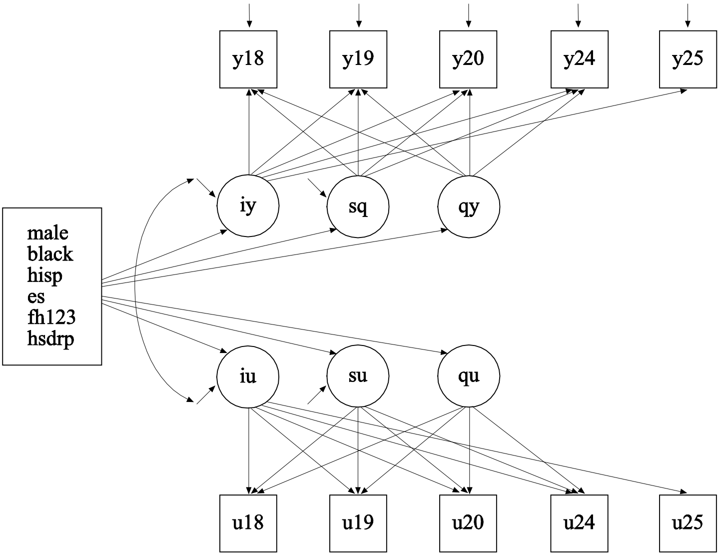
NLSY Heavy Drinking Data

- The data are from the National Longitudinal Study of Youth (NLSY), a nationally representative household study of 12,686 men and women born between 1957 and 1964.
- There are eight birth cohorts, but the current analysis considers only cohort 64 measured in 1982, 1983, 1984, 1988, 1989, and 1994 at ages 18, 19, 20, 24, and 25.
- The outcome is heavy drinking, measured by the question: How often have you had 6 or more drinks on one occasion during the last 30 days?
- The responses are coded as: never (0); once (1); 2 or 3 times (2); 4 or 5 times (3); 6 or 7 times (4); 8 or 9 times (5); and 10 or more times (6).
- Background variables include gender, ethnicity, early onset of regular drinking (es), family history of problem drinking, high school dropout and college education



hd	u	y
>0	1	log hd
0	0	999
999	999	999

NLSY Heavy Drinking Data



```
TITLE:          nlsy36425x25dep.inp
                cohort 64
                centering at 25
                hd82-hd89 (ages 18 - 25)
                log age scale: x_t = a*(ln(t-b) - ln(c-b)), where t is time, a and b
                are constants to fit the mean curve (chosen as a = 2 and b = 16),
                and c is the centering age, here set at 25.
DATA:           FILE = big.dat;
                FORMAT = 2f5, f2, t14, 5f7, t50, f8, t60, 6f1.0, t67, 2f2.0, t71,
                8f1.0, t79, f2.0, t82, 4f2.0;
DATA TWOPART:
                NAMES = hd82-hd89;
                BINARY = u18 u19 u20 u24 u25;
                CONTINUOUS = y18 y19 y20 y24 y25;
```

```
VARIABLE:   NAMES = id houseid cohort weight82 weight83 weight84
            weight88 weight89 weight94 hd82 hd83 hd84 hd88 hd89 hd94
            dep89 dep94 male black hisp es fh1 fh23 fh123 hsdrp coll ed89
            ed94 cd89 cd94;
            USEOBSERVATIONS = cohort EQ 64 AND (coll GT 0 AND
            coll LT 20);
            USEVARIABLES = male black hisp es fh123 hsdrp coll u18-
            u25 y18-y25;
            CATEGORICAL = u18-u25;
            MISSING = .;
            AUXILIARY = hd82-hd89;
DEFINE:     CUT coll (12.1);
ANALYSIS:   ESTIMATOR = ML;
            ALGORITHM = INTEGRATION;
            COVERAGE = 0.09;
```

Input For NLSY Heavy Drinking (Continued)

```
MODEL:     iu su qu | u18@-3.008 u19@-2.197 u20@-1.621 u24@-.235
           u25@.000;
           iy sy qy | y18@-3.008 y19@-2.197 y20@-1.621 y24@-.235
           y25@.000;
           iu-qy ON male black hisp es fh123 hsdrp coll;
OUTPUT:    TECH1 TECH4 TECH8 STANDARDIZED;
PLOT:      TYPE = PLOT3;
           SERIES = y18-y25(sy) | u18-u25(su);
```

Parameter	Estimate	S.E.	Est./S.E.
Regular growth modeling, treating outcome as continuous. Non-normality robust ML (MLR)			
i ON			
male	0.769	0.076	10.066
black	-0.336	0.083	-4.034
hisp	-0.227	0.103	-2.208
es	**0.291**	**0.128**	**2.283**
fh123	0.286	0.137	2.089
hsdrop	-0.024	0.104	-0.232
coll	-0.131	0.086	-1.527

Output Excerpts For Two-Part Growth Modeling Of NLSY Heavy Drinking

Parameter	Estimate	S.E.	Est./S.E.
Two-part growth modeling			
iy ON			
male	0.329	0.058	5.651
black	-0.122	0.062	-1.986
hisp	-0.143	0.069	-2.082
es	**0.096**	**0.062**	**1.543**
fh123	0.219	0.076	2.894
hsdrop	0.093	0.063	1.466
coll	-0.030	0.056	-0.526

Output Excerpts For Two-Part Growth Modeling Of NLSY Heavy Drinking

Parameter	Estimate	S.E.	Est./S.E.
iu ON			
male	1.533	0.164	5.356
black	-0.705	0.172	-4.092
hisp	-0.385	0.199	-1.934
es	**0.471**	**0.194**	**2.430**
fh123	0.287	0.224	1.281
hsdrop	-0.191	0.183	-1.045
coll	-0.325	0.161	-2.017

NLSY Heavy Drinking Conclusions

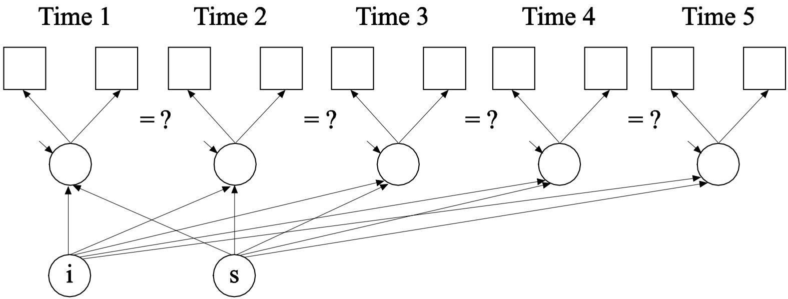
As an example of differences in results between regular growth modeling and two-part growth modeling, consider the covariate es (early start, that is, early onset of regular drinking scored as 1 if the respondent had 2 or more drinks per week at age 14 or earlier):

Regular growth modeling says that es has a significant, positive influence on heavy drinking at age 25, increasing the frequency of heavy drinking.

Two-part growth modeling says that es has a significant, positive influence on the probability of heavy drinking at age 25, but among those who engage in heavy drinking at age 25 there is no significant difference in heavy drinking frequency with respect to es.

7.8 Advances In Multiple Indicator Growth Modeling

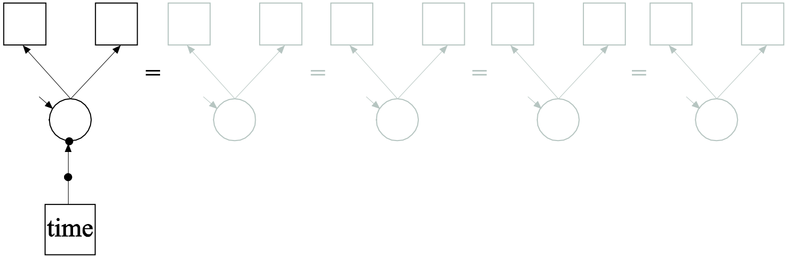
- An old dilemma
- Two new solutions



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

Categorical Items, Long Format, Two-Level Approach



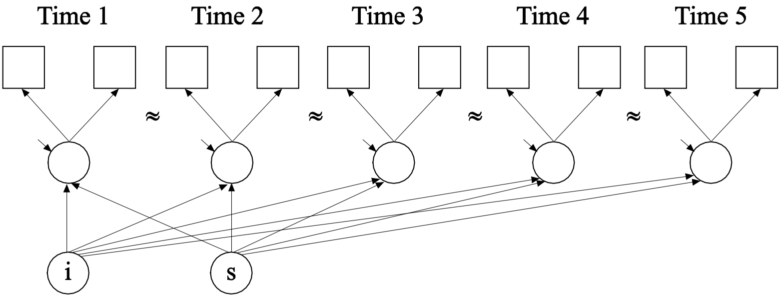
Two-level analysis with $p = 2$ variables, 1 within-factor, 2-between factors, **assuming full measurement invariance across time**.

- 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

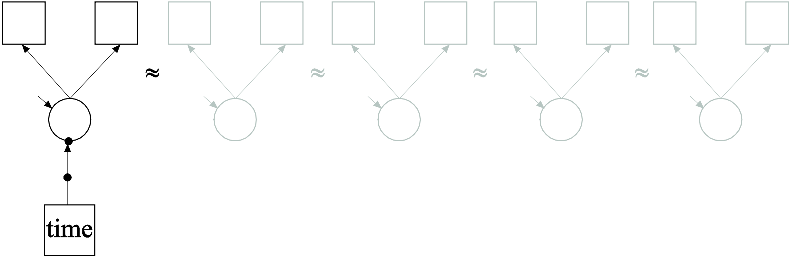
New Solution No. 1: Wide Format, Single-Level Approach



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

- 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



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

7.8.1 BSEM for 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 For Binary Items

- 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
  - 156 parameters
  - Computing time: 4:01
  - Increasing the number of time points has relatively less impact

```
VARIABLE:    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%	95% C.I. 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	*

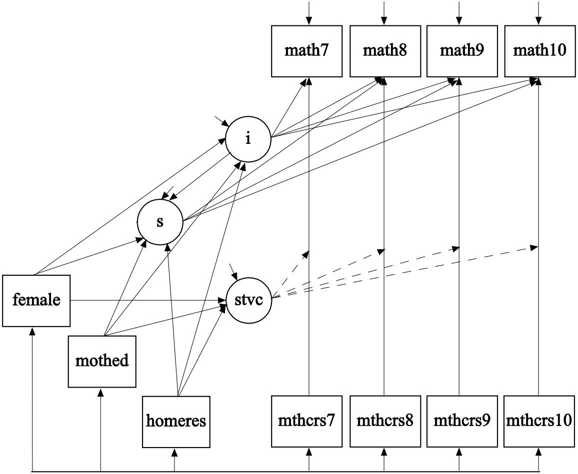
Displaying Non-Invariant Items: Time Points With Significant Differences Compared To The Mean (V = 0.01)

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

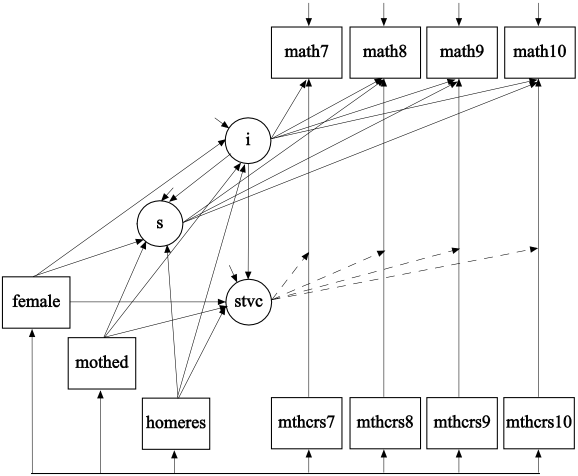
7.9 Advantages Of Growth Modeling In A Latent Variable Framework

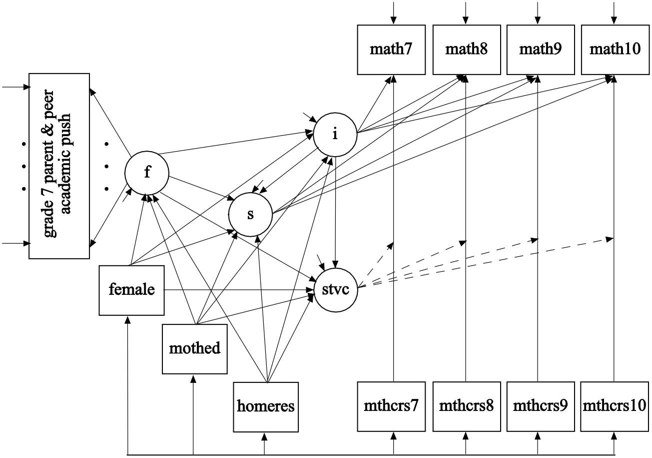
- Flexible curve shape
- Individually-varying times of observation
- Regressions among random effects
- Multiple processes
- Modeling of zeroes
- Multiple populations
- Multiple indicators
- **Embedded growth models**
- Categorical latent variables: growth mixtures

Growth Modeling With Random Slopes For Time-Varying Covariates

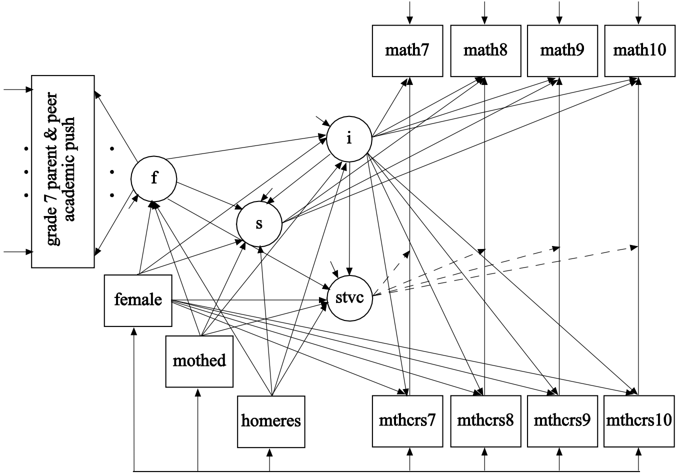


A Generalized Growth Model

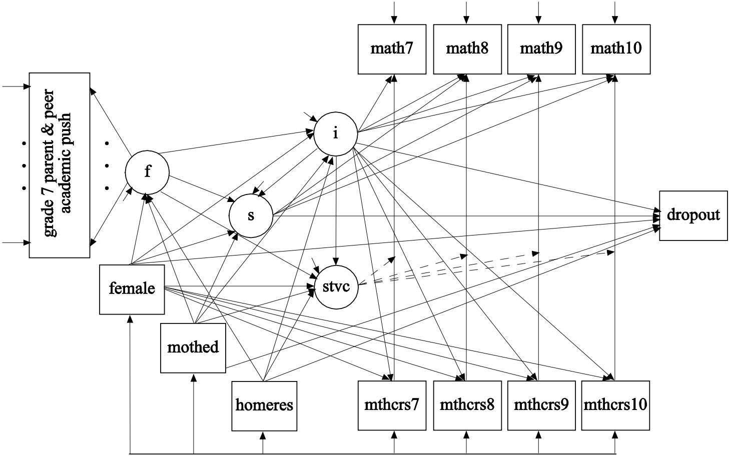



math7
math8
math9
math10
grade 7 parent & peer
academic push
f
i
s
stvc
female
mothed
homeres
mthcrs7
mthcrs8
mthcrs9
mthcrs10

A Generalized Growth Model



A Generalized Growth Model



References

For references, see handouts for Topics 1 - 9 at
`http://www.statmodel.com/course_materials.shtml`

For handouts and videos of Version 7 training, see
`http://mplus.fss.uu.nl/2012/09/12/the-workshop-new-features-of-mplus-v7/`

For papers using special Mplus features, see
`http://www.statmodel.com/papers.shtml`