

Advances in Latent Variable Modeling Using the New Mplus Version 7.2

Bengt Muthén & Tihomir Asparouhov

Mplus
www.statmodel.com

bmuthen@statmodel.com

Workshop at IMPS 2014, Madison, Wisconsin, July 21, 2014

Latent Variable Modeling in Mplus: Integration of a Multitude of Analyses

- 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

Latent Variable Modeling in Mplus: Integration of a Multitude of Analyses

- Exploratory factor analysis
- Structural equation modeling

Latent Variable Modeling in Mplus: Integration of a Multitude of Analyses

- Structural equation modeling
- Bayesian analysis

Latent Variable Modeling in Mplus: Integration of a Multitude of Analyses

- Survival analysis
- Latent class analysis

Latent Variable Modeling in Mplus: Integration of a Multitude of Analyses

- Growth mixture modeling
- Survival analysis
- Missing data modeling

Latent Variable Modeling in Mplus: Integration of a Multitude of Analyses

- Latent class analysis
- Causal inference

Latent Variable Modeling in Mplus: Integration of a Multitude of Analyses

- Structural equation modeling

- Causal inference

1. Overview of New Features in Mplus Version 7.2

- New analysis features:
 - 1 Mixture modeling with non-normal distributions: t, skew-normal, skew-t
 - 2 Structural equation modeling (SEM) with non-normal distributions: t, skew-normal, skew-t
 - 3 Mediation analysis with direct and indirect effects based on counterfactuals (causal inference)
 - 4 Latent class and latent transition analysis with residual covariances for categorical indicators
 - 5 Restructured routines for continuous-time survival analysis with latent variables
 - 6 ALIGNMENT option for binary outcomes using maximum-likelihood, including TYPE=COMPLEX using MLR
 - 7 Bootstrap standard errors and confidence intervals for maximum likelihood estimation with ALGORITHM=INTEGRATION
 - 8 Standard errors for TECH4 for the Delta parameterization of weighted least squares and z-tests and p-values for TECH4
 - 9 Standardized coefficients with standard errors for models with covariates using weighted least squares estimation

- Convenience features:
 - 1 New order of operations for the DEFINE command
 - 2 Double do loops for the DEFINE, MODEL CONSTRAINT, MODEL TEST, and MODEL PRIORS commands
 - 3 New plots: Estimated distributions; Estimated medians, modes, and percentiles; and Scatterplots of individual residuals
 - 4 For Monte Carlo studies, TYPE=TWOLEVEL, and ESTIMATOR=BAYES, the output contains a table showing the correlation and mean square error comparing true and estimated factor scores
 - 5 For the ALIGNMENT option and real data, RANKING=filename.csv; in the SAVEDATA command produces a comma-delimited file that shows the rankings of groups based on the group factor means and also shows the significance of the factor mean differences
 - 6 For Monte Carlo studies using the ALIGNMENT option, the output contains a table showing the correlation and mean square error comparing true and estimated factor means

- Convenience features, continued:
 - ① Parameter names given for parameter numbers listed as non-identified
 - ② Several lines can be commented out by starting the first line with !* and ending the last line with *!
 - ③ New features for the Mac Editor: Correction to undo function and addition of All file types
- Version 7.2 Mplus Language Addendum on the Mplus website
- Plots can now be created in R using information from most of the Mplus PLOT command options. Mplus R functions read the Mplus GH5 file using the rhdf5 package from Bioconductor, thereby providing R with the necessary input data. See <http://www.statmodel.com/mplus-R/>

2. Mediation Analysis with Effects Based on Counterfactuals/Potential Outcomes (Causal Inference; Causally-Defined Effects)

- Causal inference in mediation analysis offers counterfactually-based causal definitions of direct and indirect effects
- Research by Robins, Greenland, Pearl, VanderWeele, Vansteelandt, Imai and others
- "Little known and seldom used among SEM analysts" (Muthén & Asparouhov, 2014)

Mediation Analysis with Effects Based on Counterfactuals: Overview

- Software
- The issues, intuitively
 - Continuous Y, continuous M with "exposure-mediator interaction" ($X \times M$) influencing Y
 - Binary Y, Continuous M
- The causal effect definitions using the mediation formula
 - Specific case: Binary Y, continuous M
- Applications
 - Hopkins GBG randomized preventive intervention data
 - MacKinnon smoking data
- Sensitivity analysis (M-Y confounding)

- Focus on:
 - Binary and count Y and M
 - Single Y , single M
 - Binary (treatment/control) X or continuous (exposure) X
 - Covariates
- Valeri-VanderWeele SAS/SPSS macros (Psych Methods, 2013)
- Tingley et al. R package mediation (forthcoming in JSS)

Mplus

- 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.)
- Muthén & Asparouhov (2014). Causal effects in mediation modeling: An introduction with applications to latent variables. Forthcoming in Structural Equation Modeling
- Mplus Version 7.2 simplifies the input for the single M, single Y case
- Mplus is unique in allowing latent Y, M, and X (latent exposure), logit link without rare Y assumption, and nominal M or Y

- The effects can be estimated in Mplus using maximum-likelihood or Bayes
- ML:
 - Standard errors of the direct and indirect causal effects are obtained by the delta method using the Mplus MODEL CONSTRAINT command
 - Bootstrapped standard errors and confidence intervals are also available, taking into account possible non-normality of the effect distributions
- Bayes
 - Bayesian analysis is available in order to describe the possible non-normal posterior distributions
- Mplus Version 7.2 greatly simplifies how to get the causally-defined effects using MODEL INDIRECT (available for ML, including bootstrapping) instead of user-specified MODEL CONSTRAINT formulas

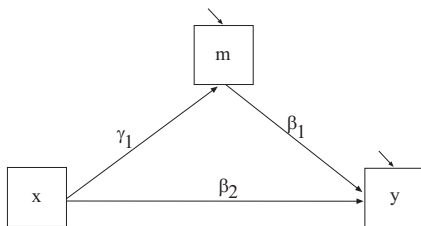
Mplus Version 7.2 MODEL INDIRECT Commands for Causal Effects

- 1 No moderation:
 - Y IND M X;
 - - all 3 can be latent
- 2 Moderation with X*M:
 - Y MOD M XM X;
 - Y can be latent
- 3 Moderation with Z involving X and M:
 - Y MOD M Z(low, high, increment) MZ XZ X;
 - - only Y can be latent
- 4 Moderation with Z involving M and not X:
 - Y MOD M Z(low, high, increment) MZ X;
 - - X and Y can be latent
- 5 Moderation with Z involving X and not M:
 - Y MOD M Z(low, high, increment) XZ X;
 - - M and Y can be latent

For controlled direct effects an M value is placed in parenthesis:
M(m).

2.2 The Issues, Intuitively

- Causally-defined effects based on counterfactuals and potential outcomes using expectations have been developed by Robins, Greenland, Pearl, VanderWeele, Vansteelandt, Imai etc
 - Total, direct, and indirect causal effects
- Different results than SEM with for instance "exposure-mediator interaction" ($Y=X*M$), categorical DV, or count DV
- The effects are causal only under strong assumptions (if assumptions don't hold, are the causal methods better/useful anyway?)



Valeri and VanderWeele (2011):

”(i) no unmeasured confounding of the treatment-outcome relationship.

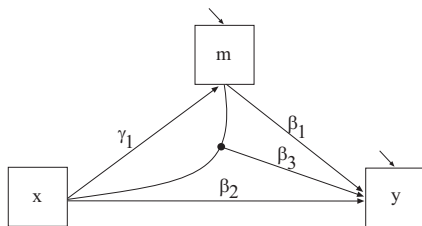
(ii) no unmeasured confounding of the mediator-outcome relationship.

(iii) no unmeasured treatment-mediator confounding

(iv) no mediator-outcome confounder affected by treatment”

Assumptions (i) and (iii) fulfilled when treatment X is randomized.

Continuous Y and M with Exposure-Mediator Interaction



$$\text{Total natural indirect effect : } TNIE = \beta_1 \gamma_1 + \beta_3 \gamma_1. \quad (1)$$

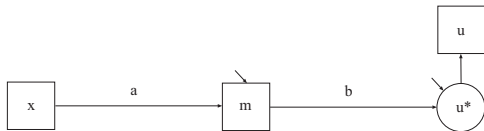
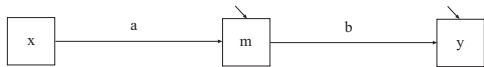
$$\text{Pure natural indirect effect : } PNIE = \beta_1 \gamma_1. \quad (2)$$

$$\text{Pure natural direct effect : } PNDE = \beta_2 + \beta_3 \gamma_0. \quad (3)$$

$$\text{Total natural direct effect : } TNDE = \beta_2 + \beta_3 \gamma_0 + \beta_3 \gamma_1. \quad (4)$$

$$\text{Total effect} = (1) + (3) = (2) + (4)$$

Continuous versus Binary Distal Outcome



Conventional versus Causal Mediation Effects with a Categorical Distal Outcome

With a categorical distal outcome, the conventional product formula for an indirect effect is only valid for an underlying continuous latent response variable behind the categorical observed outcome (2 linear regressions), not for the observed categorical outcome itself (linear plus non-linear regression).

Similarly, with a categorical mediator, conventional product formulas for indirect effects are only relevant/valid for a continuous latent response variable behind the mediator.

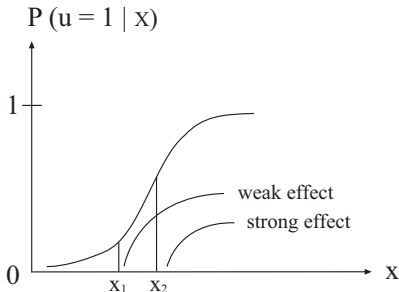
The difference between the causal effects and the effects obtained by what is called the naive approach has been studied in Imai et al. (2010a) and Pearl (2011c). Imai et al. (2010a, Appendix E, p. 23) conducted a Monte Carlo simulation study to show the biases, while Pearl (2011c) presented graphs showing the differences.

We will look at the differences in some examples.

Previous use of the Product Approach with Binary Distal Outcome

- MacKinnon & Dwyer (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17, 144-158
- MacKinnon, D.P., Lockwood, C.M., Brown, C.H., Wang, W., & Hoffman, J.M. (2007). The intermediate endpoint effect in logistic and probit regression. *Clinical Trials*, 4, 499-513
- Mplus MODEL INDIRECT

The Problem with $a \times b$ for a Binary Outcome: 2 Parameters used when 5 are Needed



- The $a \times b$ indirect effect faces the problem of non-constant effect due to ignoring the level parameters (the intercept for the mediator and threshold for the distal outcome)
- The causally-defined indirect effect uses these level parameters, focusing on the expected values of the observed binary outcome - the probabilities

The Probability of $u = 1|x$

- Conditional on x , $P(u = 1|x)$ is obtained by integrating over the residual of the mediator and apart from the regression coefficients the probability involves the residual variance, the mediator intercept, and the distal outcome threshold: 5 parameters
- One can compute $P(u = 1|x = 1) - P(u = 1|x = 0)$ to compare treatment and control
- This is TE (Total Effect)
- But what are the indirect and direct effects?
- This is where the counterfactual definitions of causal effects come in

2.3 Causal Effect Definitions

- $Y_i(x)$: Potential outcome that would have been observed for that subject had the treatment variable X been set at the value x , where x is 0 or 1 in the example considered here
- The $Y_i(x)$ outcome may not be the outcome that is observed for the subject and is therefore possibly counterfactual
- The causal effect of treatment for a subject can be seen as $Y_i(1) - Y_i(0)$, but is clearly not identified given that a subject only experiences one of the two treatments
- The average effect $E[Y(1) - Y(0)]$ is, however, identifiable
- Similarly, let $Y(x, m)$ denote the potential outcome that would have been observed if the treatment for the subject was x and the value of the mediator M was m

The controlled direct effect is defined as (C is a covariate)

$$CDE(m) = E[Y(1, m) - Y(0, m) \mid C = c]. \quad (5)$$

where $M = m$ for a fixed value m . The first index of the first term is 1 corresponding to the treatment group and the first index of the second term is 0 corresponding to the control group.

VanderWeele-Vansteelandt (2009):

While controlled direct effects are often of greater interest in policy evaluation (Pearl, 2001; Robins, 2003), natural direct and indirect effects may be of greater interest in evaluating the action of various mechanisms (Robins, 2003; Joffe et al., 2007).

CDE is the same as direct effect for regular mediation model with continuous Y , M and no $X \times M$ interaction.

The Pure Natural Direct Effect (PNDE)

The pure natural direct effect does not hold the mediator constant, but instead allows the mediator to vary over subjects in the way it would vary if the subjects were given the control condition ("effect on the mediator somehow blocked"). The direct effect is expressed as

$$PNDE = E[Y(1, M(0)) - Y(0, M(0)) | C = c] = \quad (6)$$

$$= \int_{-\infty}^{\infty} \{E[Y | C = c, X = 1, M = m] - E[Y | C = c, X = 0, M = m]\} \\ \times f(M | C = c, X = 0) \partial M, \quad (7)$$

where f is the density of M . A simple way to view this is to note that Y 's first argument, that is x , changes values, but the second does not, implying that Y is influenced by X only directly. The right-hand side of (7) is part of what is referred to as the Mediation Formula in Pearl (2009, 2011c).

The total natural indirect effect is defined as (Robins, 2003)

$$TNIE = E[Y(1, M(1)) - Y(1, M(0)) | C = c] = \quad (8)$$

$$= \int_{-\infty}^{\infty} E[Y | C = c, X = 1, M = m] \times f(M | C = c, X = 1) \partial M$$

$$- \int_{-\infty}^{\infty} E[Y | C = c, X = 0, M = m] \times f(M | C = c, X = 0) \partial M. \quad (9)$$

A simple way to view this is to note that the first argument of Y does not change, but the second does, implying that Y is influenced by X due to its influence on M .

The total effect is (Robins, 2003)

$$TE = E[Y(1) - Y(0) \mid C = c] \quad (10)$$

$$= E[Y(1, M(1)) - Y(0, M(0)) \mid C = c]. \quad (11)$$

A simple way to view this is to note that both indices are 1 in the first term and 0 in the second term. In other words, the treatment effect on Y comes both directly and indirectly due to M . The total effect is the sum of the direct effect and the total indirect effect (Robins, 2003),

$$TE = PNDE + TNIE. \quad (12)$$

The pure natural indirect effect (Robins, 2003) is defined as

$$PNIE = E[Y(0, M(1)) - Y(0, M(0)) | C = c] \quad (13)$$

Here, the effect of X on Y is only indirect via M. This is called the natural indirect effect in Pearl (2001) and VanderWeele and Vansteelandt (2009).

Translation of Different Terms by Different Authors

Expectation Setting	Names of Effects		
	Imai	Pearl/VanderWeele	Robins
1,1 - 1,0	ACME(treated)	TNIE	TIE
0,1 - 0,0	ACME(control)	PNIE	PIE
1,0 - 0,0	ADE(control)	PNDE	DE
1,1 - 0,1	ADE(treated)	TNDE	TDE
1,1 - 0,0	Total	Total	Total

ACME - Average causal mediated effect

ADE - Average direct effect

TNIE - Total natural indirect effect

PNIE - Pure natural indirect effect

PNDE - Pure natural direct effect

TNDE - Total natural direct effect

TIE - Total indirect effect

PIE - Pure indirect effect

DE - direct effect

Natural (N) direct (D) and indirect (I) effect decompositions of the total effect (TE) can be expressed in two ways:

① $TE = \text{Pure NDE} + \text{Total NIE} = \text{PNDE} + \text{TNIE} (= \text{DE} + \text{TIE})$

- $E[Y(1, M(1)) - Y(0, M(0))] =$
 $E[Y(1, M(0)) - Y(0, M(0))] + E[Y(1, M(1)) - Y(1, M(0))]$

② $TE = \text{Total NDE} + \text{Pure NIE} = \text{TNDE} + \text{PNIE}$

- $E[Y(1, M(1)) - Y(0, M(0))] =$
 $E[Y(1, M(1)) - Y(0, M(1))] + E[Y(0, M(1)) - Y(0, M(0))]$

① is the focus of Valeri-VanderWeele (2013).

The causal effects are expressed in a general way using expectations and can be applied to many different settings:

- Continuous mediator, continuous distal outcome (gives the usual SEM formulas)
- Categorical mediator, continuous distal outcome
- Continuous mediator, categorical distal outcome
- Categorical mediator, categorical distal outcome
- Count distal outcome
- Nominal mediator, nominal outcome
- Survival distal outcome

Using the general definition, the causal natural indirect effect (total indirect effect) is expressed as the probability difference

$$\text{Total NIE} = \text{TIE} = \Phi[\text{probit}(1, 1)] - \Phi[\text{probit}(1, 0)], \quad (14)$$

where Φ is the standard normal distribution function, the argument $(a, b) = (x, M(x))$, and *probit* is defined on the next slide.

The pure natural indirect effect is expressed as the probability difference

$$\text{Pure NIE} = \text{PIE} = \Phi[\text{probit}(0, 1)] - \Phi[\text{probit}(0, 0)]. \quad (15)$$

and the pure natural direct effect expressed as the probability difference

$$\text{Pure NDE} = \text{DE} = \Phi[\text{probit}(1, 0)] - \Phi[\text{probit}(0, 0)], \quad (16)$$

$$\text{TE} = \text{Pure NDE} + \text{Total NIE} = \text{DE} + \text{TIE} = \Phi[\text{probit}(1, 1)] - \Phi[\text{probit}(0, 0)]. \quad (17)$$

Binary Distal Outcome Continued

Consider a mediation model for a binary outcome u and a continuous mediator m . Assume a probit link for the binary outcome u ,

$$\text{probit}(u_i) = \beta_0 + \beta_1 m_i + \beta_2 x_i + \beta_3 x_i m_i + \beta_4 c_i, \quad (18)$$

$$m_i = \gamma_0 + \gamma_1 x_i + \gamma_2 c_i + \varepsilon_{2i}, \quad (19)$$

where the residual ε_2 is assumed normally distributed. For $x, x' = 0, 1$ corresponding to the control and treatment group,

$$\text{probit}(x, x') = [\beta_0 + \beta_2 x + \beta_4 c + (\beta_1 + \beta_3 x)(\gamma_0 + \gamma_1 x' + \gamma_2 c)] / \sqrt{v(x)}, \quad (20)$$

where the variance $v(x)$ for $x = 0, 1$ is

$$v(x) = (\beta_1 + \beta_3 x)^2 \sigma_2^2 + 1. \quad (21)$$

where σ_2^2 is the residual variance for the continuous mediator m . Although not expressed in simple functions of model parameters, the quantity of (14) can be computed and corresponds to the change in the $y=1$ probability due to the indirect effect of the treatment (conditionally on c when that covariate is present).

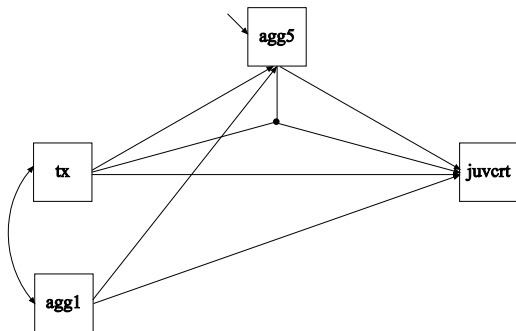
2.4 Example: Aggressive Behavior and Juvenile Court Record

- Randomized field experiment in Baltimore public schools
- Classroom-based intervention aimed at reducing aggressive-disruptive behavior among elementary school students
- Covariate is the Grade 1 aggression score before the intervention started
- Mediator is the aggression score in Grade 5 after the intervention ended
- Distal outcome is a binary variable indicating whether or not the student obtained a juvenile court record by age 18 or an adult criminal record
- $n = 250$ boys in treatment and control classrooms

A Mediation Model for Aggressive Behavior and a Binary Juvenile Court Outcome

Two reasons for causal effects:

- "Exposure-mediator interaction" ($tx*agg5$)
- Binary outcome ($juvcrt$)



A Mediation Model for Aggressive Behavior and a Binary Juvenile Court Outcome

$$juvcrt_i^* = \beta_0 + \beta_1 agg5_i + \beta_2 tx_i + \beta_3 tx_i agg5_i + \beta_4 agg1_i + \varepsilon_{1i}, \quad (22)$$

$$agg5_i = \gamma_0 + \gamma_1 tx_i + \gamma_2 agg1_i + \varepsilon_{2i}. \quad (23)$$

The *juvcrt* outcome is not rare, but is observed for 50% of the sample. The mediator *agg5* is not normally distributed, but is quite skewed with a heavy concentration at low values. The normality assumption, however, pertains to the mediator residual ε_2 . Because the covariate *agg1* has a distribution similar to the mediator *agg5*, the *agg5* distribution is to some extent produced by the *agg1* distribution so that the normality assumption for the residual may be a reasonable approximation.

Aggressive Behavior and Juvenile Court Record: Mplus Input for Causal Effects

Analysis:

```
estimator = mlr;  
link = probit;  
integration = montecarlo;
```

model:

```
[juvcrt$1] (mbeta0);  
juvcrt on tx (beta2)  
agg5 (beta1)  
xm (beta3)  
agg1 (beta4);  
[agg5] (gamma0);  
agg5 on tx (gamma1)  
agg1 (gamma2);  
agg5 (sig2);
```

Aggressive Behavior and Juvenile Court Record: Mplus Input for Causal Effects, Continued - Done Automatically in Mplus Version 7.2

model constraint:

```
new(ind dir arg11 arg10 arg00 v1 v0
probit11 probit10 probit00 indirect direct
total letete complete ori nd ordi r);
dir=beta3*gamma0+beta2;
ind=beta1*gamma1+beta3*gamma1;
arg11=-mbeta0+beta2+beta4*0+(beta1+beta3)*(gamma0+gamma1+gamma2*0);
arg10=-mbeta0+beta2+(beta1+beta3)*gamma0;
arg00=-mbeta0+beta1*gamma0;
v1=(beta1+beta3)^2*sig2+1;
v0=beta1^2*sig2+1;
probit11=arg11/sqrt(v1);
probit10=arg10/sqrt(v1);
probit00=arg00/sqrt(v0);
! Version 6.12 Phi function needed below:
indirect=phi(probit11)-phi(probit10);
direct=phi(probit10)-phi(probit00);
total=phi(probit11)-phi(probit00);
ori nd=(phi(probit11)/(1-phi(probit11)))/(phi(probit10)/(1-phi(probit10)));
ordi r=(phi(probit10)/(1-phi(probit10)))/(phi(probit00)/(1-phi(probit00)));
```

Instead of MODEL CONSTRAINT, use slide 10 language, case 2:
MODEL INDIRECT:

juvcrt MOD agg5 xm tx;

TOTAL, INDIRECT, AND DIRECT EFFECTS BASED ON
COUNTERFACTUALS (CAUSALLY-DEFINED EFFECTS)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Effects from TX to JUVCRT				
Pure natural DE	0.005	0.067	0.076	0.940
Tot natural IE	-0.064	0.030	-2.158	0.031
Total effect	-0.059	0.070	-0.848	0.397
Odds ratios for binary Y				
Pure natural DE	1.021	0.275	3.714	0.000
Tot natural IE	0.773	0.092	8.371	0.000
Total effect	0.789	0.221	3.571	0.000
Other effects				
Tot natural DE	-0.015	0.066	-0.219	0.826
Pure natural IE	-0.045	0.022	-2.056	0.040
Total effect	-0.059	0.070	-0.848	0.397
Odds ratios for other effects for binary Y				
Tot natural DE	0.943	0.251	3.760	0.000
Pure natural IE	0.837	0.073	11.507	0.000
Total effect	0.789	0.221	3.571	0.000

The causal direct effect is not significant. The causal indirect effect is estimated as -0.064 and is significant ($z = -2.120$). This is the drop in the probability of a juvenile court record due to the indirect effect of treatment.

The odds ratio for the indirect effect is estimated as 0.773 which is significantly different from one ($z = (0.773 - 1)/0.092 = -2.467$).

The conventional direct effect is not significant and the conventional product indirect effect is -0.191 ($z = -1.98$).

Mediated Moderation Using Mplus Version 7.2

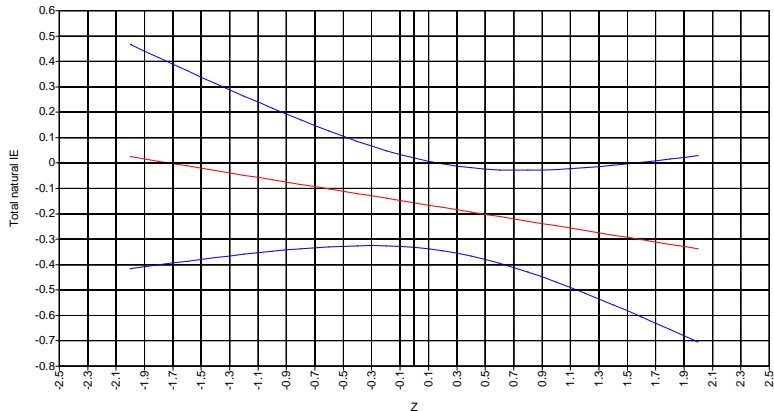
MODEL INDIRECT and LOOP Plot for Moderator Z

Excerpts from the Version 7.2 Mplus Language Addendum, pp. 3-8.
Language follows slide 10, Case 5.

The following input shows how the new MODEL INDIRECT language can be used as an alternative to the use of the PLOT and LOOP options of the MODEL CONSTRAINT command to get estimates and plots of the moderated indirect effect as shown in Example 3.18. Instead of the Bayesian analysis of Example 3.18, maximum-likelihood estimation is used with standard errors and confidence intervals obtained by bootstrapping.

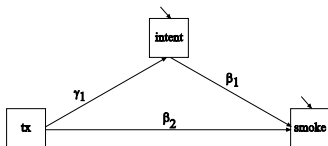
```
TITLE:      this is an example of moderated mediation with a
            plot of the indirect effect as in Example 3.18
            but using bootstrap and maximum likelihood
            estimation
DATA:      FILE = ex3.18.dat;
VARIABLE:  NAMES = y m x z;
            USEVARIABLES = y m x z xz;
DEFINE:    xz = x*z;
ANALYSIS:  BOOTSTRAP = 500;
MODEL:     y ON m xz z;
            m ON z xz x;
MODEL INDIRECT:
            y MOD m z (-2 2 0.1) xz x;
PLOT:     TYPE = PLOT2;
OUTPUT:    CINTERVAL (BCBOOTSTRAP);
```

LOOP Plot for Moderated Mediation: Indirect Effect as a Function of the Moderator Z



3.5 Categorical Mediator: Smoking Data Example

Muthén (2011): MacKinnon et al (2007) smoking data (binary Y)



	Cigarette use			
	Intention	No Use	Use	Total
Ctrl	4 (Yes)	9	20	29
	3 (Probably)	14	20	34
	2 (Don't think so)	36	13	49
	1 (No)	229	30	259
Tx	4 (Yes)	9	19	28
	3 (Probably)	15	11	26
	2 (Don't think so)	43	11	54
	1 (No)	353	32	385

Different approaches with an ordinal mediator:

- Ordered polytomous variable treated as ordered polytomous variable
- Ordered polytomous variable treated as continuous (non-normal residual issue)
- Latent response variable behind ordered polytomous variable
- Dichotomized variable
- Latent response variable behind dichotomized variable

See Muthén (2011)

Latent response variable behind categorical variable available with WLSMV and Bayes, but not ML.

3.6 Binary Mediator and Binary Distal Outcome

Recalling that the general formulas for the direct, total indirect, and pure indirect effects are defined as

$$PNDE = E[Y(1, M(0)) - Y(0, M(0)) | C], \quad (24)$$

$$TNIE = E[Y(1, M(1)) - Y(1, M(0)) | C], \quad (25)$$

$$PNIE = E[Y(0, M(1)) - Y(0, M(0)) | C], \quad (26)$$

it can be shown that with a binary mediator and a binary outcome these formulas lead to the expressions

$$PNDE = [F_Y(1, 0) - F_Y(0, 0)] [1 - F_M(0)] + [F_Y(1, 1) - F_Y(0, 1)] F_M(0), \quad (27)$$

$$TNIE = [F_Y(1, 1) - F_Y(1, 0)] [F_M(1) - F_m(0)], \quad (28)$$

$$PNIE = [F_Y(0, 1) - F_Y(0, 0)] [F_M(1) - F_m(0)]. \quad (29)$$

where $F_Y(x, m)$ denotes $P(Y = 1 | X = x, M = m)$ and $F_M(x)$ denotes $P(M = 1 | X = x)$, where F denotes either the standard normal or the logistic distribution function corresponding to using probit or logistic regression. These formulas agree with those of Pearl (2010, 2011a).

Pearl (2010, 2011a) provided a hypothetical example with a binary treatment X , a binary mediator M corresponding to the enzyme level in the subject's blood stream, and a binary outcome Y corresponding to being cured or not. This example was also hotly debated on SEMNET in September 2011.

Pearl's Hypothetical Binary-Binary Case, Continued

Treatment X	Enzyme M	Percentage Cured Y = 1
1	1	$F_Y(1, 1) = 80\%$
1	0	$F_Y(1, 0) = 40\%$
0	1	$F_Y(0, 1) = 30\%$
0	0	$F_Y(0, 0) = 20\%$

Treatment	Percentage M=1
0	$F_M(0) = 40\%$
1	$F_M(1) = 75\%$

The top part of the table suggests that the percentage cured is higher in the treatment group for both enzyme levels and that the effect of treatment is higher at enzyme level 1 than enzyme level 0:
Treatment-mediator interaction.

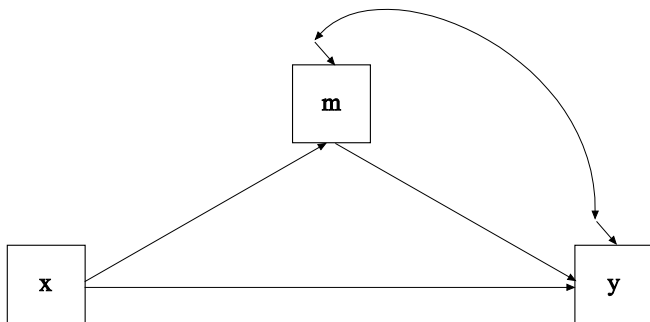
2.7 Sensitivity Analysis of Mediator-Outcome Confounding

To claim that effects are causal, it is not sufficient to simply use the causally-derived effects

The underlying assumptions need to be fulfilled, such as no mediator-outcome confounding

Violation of the no mediator-outcome confounding can be seen as an unmeasured (latent) variable Z influencing both the mediator M and the outcome Y . When Z is not included in the model, a covariance is created between the residuals in the two equations of the regular mediation model. Including the residual covariance, however, makes the model not identified.

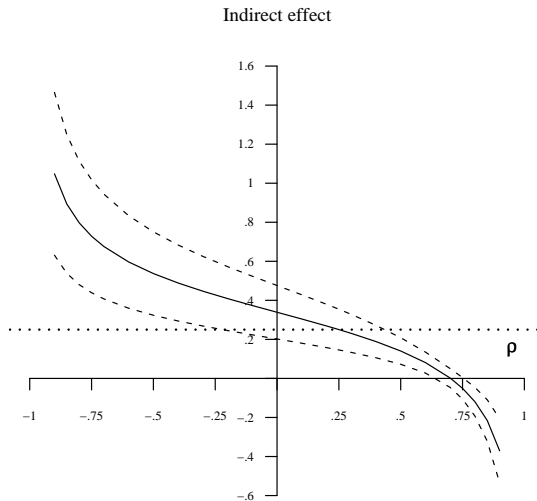
Mediator-Outcome Confounding (Residual Correlation $\rho \neq 0$)



Imai et al. (2010a, b) proposed a sensitivity analysis where causal effects are computed given different fixed values of the residual covariance. This is useful both in real-data analyses as well as in planning studies. As for the latter, the approach can answer questions such as how large does your sample and effects have to be for the lower confidence band on the indirect effect to not include zero when allowing for a certain degree of mediator-outcome confounding?

Sensitivity plots can be made in Mplus using LOOP in the PLOT command.

Indirect Effect Based on Imai Sensitivity Analysis with ρ Varying from -0.9 to +0.9 and True Residual Correlation 0.25



Explaining the Sensitivity Figure

- The correct value for the indirect effect is 0.25 (marked with a horizontal broken line)
- The biased estimate assuming $\rho = 0$ is 0.3287, an overestimation due to ignoring the positive residual correlation
- The sensitivity analysis varies the ρ values from -0.9 to $+0.9$:
 - Using $\rho = 0$, the biased estimate of 0.3287 is obtained
 - Using the correct value of $\rho = 0.25$, the correct indirect effect value of 0.25 is obtained
 - For lower ρ values the effect is overestimated and for larger ρ values the effect is underestimated
- The graph provides useful information for planning new studies:
 - At this sample size ($n = 400$) and effect size, the lower confidence limit does not include zero until about $\rho = 0.6$
 - This means that a rather high degree of confounding is needed for the effect to not be detected
 - In the range of ρ from about -0.1 to $+0.4$ the confidence interval covers the correct value of 0.25 for the indirect effect

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
- Non-informative versus informative priors
- 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
- Informative priors can better reflect substantive hypotheses
- Analyses can be made less computationally demanding
- New types of models can be analyzed

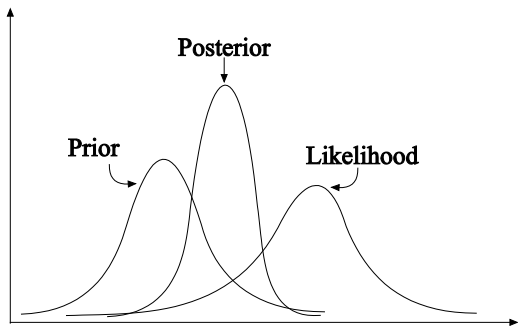
For a Bayes introduction with further references, see, e.g., Muthén (2010). Bayesian analysis in Mplus: A brief introduction. Technical Report. Version 3.

- Muthén (2010). Bayesian analysis in Mplus: A brief introduction. Technical Report. Version 3.
- 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 & 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.
- Asparouhov & Muthén (2012). General random effect latent variable modeling: Random subjects, items, contexts, and parameters.
- Asparouhov & Muthén (2012). Comparison of computational methods for high dimensional item factor 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



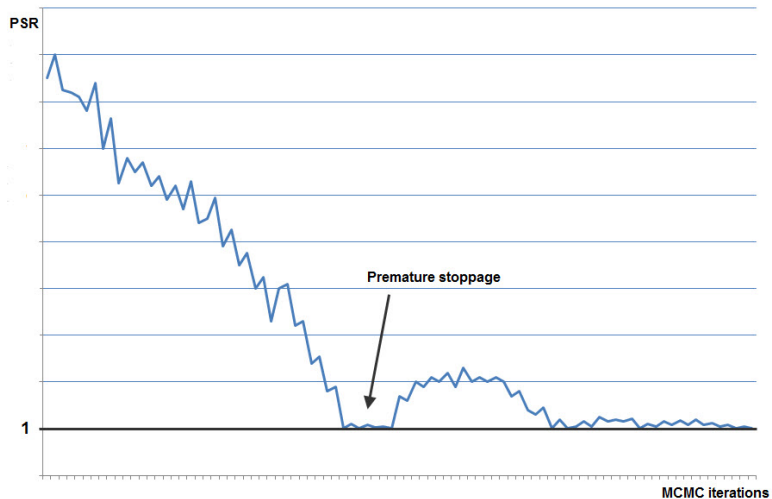
Bayesian Estimation Obtained Iteratively Using Markov Chain Monte Carlo (MCMC) Algorithms

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

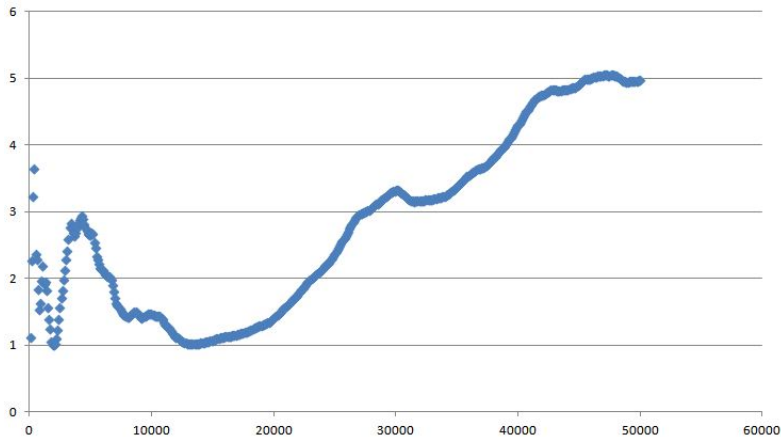
Asparouhov & Muthén (2010). Bayesian analysis using Mplus.
Technical implementation. Technical Report.

- 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 $PSR \approx 1$

PSR Convergence Issues: Premature Stoppage



PSR Convergence Issues: Premature Stoppages Due to Non-Identification



- Topic 9 handout and video from the Johns Hopkins teaching, June 2011
- Part 1 - 3 handouts and videos from the Utrecht University teaching, August 2012
- Papers on our website under Bayesian Analysis, such as
 - Muthén (2010). Bayesian analysis in Mplus: A brief introduction. Technical Report
 - Muthén & Asparouhov (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, 17, 313-335.

Types of factor analyses in Mplus:

- EFA: Exploratory Factor Analysis - Regular and bi-factor rotations
- CFA: Confirmatory Factor Analysis
- ESEM: Exploratory Structural Equation Modeling (Asparouhov & Muthén, 2009 in Structural Equation Modeling)
- BSEM: Bayesian Structural Equation Modeling (Muthén & Asparouhov, 2012 in Psychological Methods)
- skew-SEM: Structural equation modeling with continuous non-normal skewed distributions (Asparouhov & Muthén, 2014 forthcoming in Structural Equation Modeling)

Consider the MIMIC model

$$Y = \nu + \lambda\eta + \beta X + \varepsilon$$

$$\eta = \alpha + \gamma X + \zeta$$

Mplus can be used to compute the factor scores $\hat{\eta} = E(\eta|Y, X)$ and the model estimated/predicted values for Y, which we denote by \hat{Y} . Two versions of \hat{Y} ,

$$\hat{Y}_a = \hat{\nu} + \hat{\lambda}\hat{\eta} + \hat{\beta}X,$$

$$\hat{Y}_b = E(\hat{Y}|X).$$

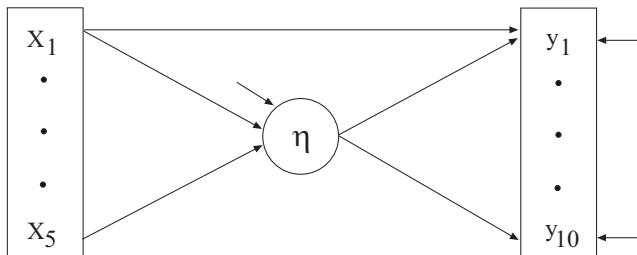
Individual level residuals can then be formed as

$$Y_{res,a} = Y - \hat{Y}_a$$

$$Y_{res,b} = Y - \hat{Y}_b.$$

Source: Asparouhov & Muthén (2013). Using Mplus plots for diagnosis and model evaluation in SEM.

When inadvertently left out, can the need for a direct effect be seen in the individual-specific residuals?



$Y_{res,b} = Y - \hat{Y}_b$ is useful for this.

Figure : $\hat{Y}_{1,res,b}$ v.s. X_1

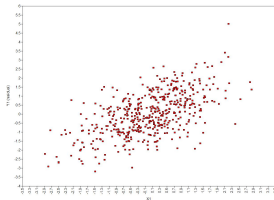
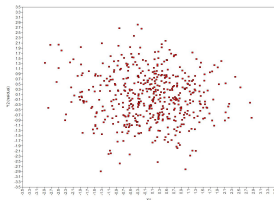


Figure : $\hat{Y}_{2,res,b}$ v.s. X_1



- 2-parameter logit and probit
- Weighted least-squares, maximum-likelihood, and Bayes estimators for multidimensional analysis (Asparouhov & Muthén, 2012a)
- Two-level, three-level, and cross-classified analysis (Asparouhov & Muthén, 2012b)
- Mixture modeling (Muthén, 2008; Muthén & Asparouhov, 2009)
- Multilevel mixture modeling (Asparouhov & Muthén, 2008; Henry & Muthén, 2010)
- Alignment approach to the study of measurement invariance with many groups (Asparouhov & Muthén, 2013; Muthén & Asparouhov, 2013)

Table : Comparisons of estimators for categorical factor analysis (+ implies an advantage and - implies a disadvantage)

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

5. Analysis Choices for Multiple Groups/Clusters: Fixed vs Random Effect Factor Analysis (IRT)

- Fixed mode: Multiple-group factor analysis
 - Inference to the groups in the sample
 - Usually a relatively small number of groups
- Random mode: Two-level factor analysis
 - Inference to a population from which the groups/clusters have been sampled
 - Usually a relatively large number of groups/clusters

5.1 Papers on New Techniques

- Fixed mode:
 - ESEM:
 - Asparouhov & Muthén (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397-438
 - Alignment:
 - Asparouhov & Muthén (2013). Multiple group factor analysis alignment. Web note 18. Forthcoming in SEM
 - Muthén & Asparouhov (2013). New methods for the study of measurement invariance with many groups
 - Muthén & Asparouhov (2014). IRT studies of many groups: The alignment method. Forthcoming in *Frontiers of Psychology*
 - BSEM:
 - Muthén & Asparouhov (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*
 - Muthén & Asparouhov (2013). BSEM measurement invariance analysis. Web note 17
- Random mode:
 - Two-level (random intercepts and loadings):
 - Fox (2010). Bayesian IRT.
 - Asparouhov & Muthén (2012). General random effect latent variable modeling: Random subjects, items, contexts, parameters.

5.2 Fixed Mode, Traditional, Exact Invariance Modeling: Refresher on Multiple-Group Factor Analysis: 3 Different Degrees of Measurement Invariance

- 1 CONFIGURAL (invariant factor loading pattern)
- 2 METRIC (invariant factor loadings; "weak factorial invariance")
 - Needed in order to compare factor variances across groups
- 3 SCALAR (invariant factor loadings and intercepts/thresholds; "strong factorial invariance")
 - Needed in order to compare factor means across groups

Refresher on Multiple-Group Factor Analysis: Formulas for Individual i and Group j

- Configural:

$$y_{ij} = \nu_j + \lambda_j f_{ij} + \varepsilon_{ij},$$
$$E(f_j) = \alpha_j = 0, V(f_j) = \psi_j = 1.$$

- Metric:

$$y_{ij} = \nu_j + \lambda f_{ij} + \varepsilon_{ij},$$
$$E(f_j) = \alpha_j = 0, V(f_j) = \psi_j.$$

- Scalar:

$$y_{ij} = \nu + \lambda f_{ij} + \varepsilon_{ij},$$
$$E(f_j) = \alpha_j, V(f_j) = \psi_j.$$

Measurement invariance ("item bias", "DIF") has traditionally been concerned with comparing a small number of groups such as with gender or ethnicity.

Likelihood-ratio chi-square testing of one item at a time:

- Bottom-up: Start with no invariance (configural case), imposing invariance one item at a time
- Top-down: Start with full invariance (scalar case), freeing invariance one item at a time, e.g. using modification indices

Neither approach is scalable - both are very cumbersome when there are many groups, such as 50 countries ($50 \times 49/2 = 1225$ pairwise comparisons for each item). The correct model may well be far from either of the two starting points, which may lead to the wrong model.

5.3 Multiple-Group Factor Analysis: A New Method - Alignment Optimization

There is a need for a new approach to multiple-group factor analysis for many groups such as with country comparisons of achievement (PISA, TIMSS, PIRL) or cross-cultural studies (ISSP, ESS etc):

- Goal is to study measurement invariance and also group differences in factor means and variances
- Standard approach is confirmatory factor analysis with equality constraints, followed by model modifications
- The standard approach is too cumbersome to be practical for analysis of many groups where there can be a large number of non-invariant measurement parameters
- A radically different method was introduced in Mplus Version 7.1: Alignment optimization
- Alignment was expanded in Mplus Version 7.2 to ML estimation with binary items and complex survey data

Multiple-Group CFA Alignment Optimization

- 1 Estimate the configural model (loadings and intercepts free across groups, factor means fixed @0, factor variances fixed @1)
- 2 Alignment optimization:
 - Free the factor means and variances and choose their values to minimize the total amount of non-invariance using a simplicity function

$$F = \sum_P \sum_{j_1 < j_2} w_{j_1, j_2} f(\lambda_{pj_1} - \lambda_{pj_2}) + \sum_P \sum_{j_1 < j_2} w_{j_1, j_2} f(v_{pj_1} - v_{pj_2}),$$

for every pair of groups and every intercept and loading using a component loss function (CLF) f from EFA rotations (Jennrich, 2006)

- The simplicity function F is optimized at a few large non-invariant parameters and many approximately invariant parameters rather than many medium-sized non-invariant parameters (compare with EFA rotations using functions that aim for either large or small loadings, not mid-sized loadings)

- In this way, a non-identified model where factor means and factor variances are added to the configural model is made identified by adding a simplicity requirement
- This model has the same fit as the configural model:
 - Free the factor means α_j and variances ψ_j , noting that for every set of factor means and variances the same fit as the configural model is obtained with loadings λ_j and intercepts ν_j changed as:

$$\lambda_j = \lambda_{j,\text{configural}} / \sqrt{\psi_j},$$

$$\nu_j = \nu_{j,\text{configural}} - \alpha_j \lambda_{j,\text{configural}} / \sqrt{\psi_j}.$$

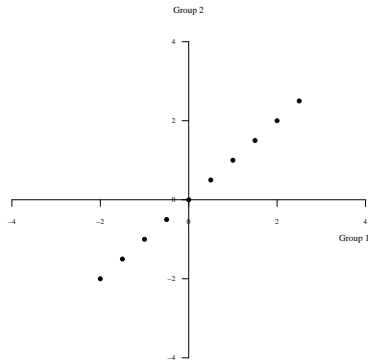
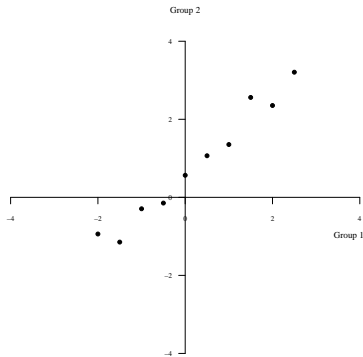
- Simulation studies show that the alignment method works very well unless there is a majority of significant non-invariant parameters or small group sizes
- For well-known examples with few groups and few non-invariances, the results agree with the alignment method

A Visual Answer to Why it is Called Alignment

Consider group-invariant intercepts for 10 items and 2 groups with factor means = 0, -1 and factor variances = 1, 2

Unaligned: Configural model
(mean=0, variance=1 in both groups)

Aligned: Taking into account the
group differences in means and
variances



In addition to the estimated aligned model, the alignment procedure gives

- Measurement invariance test results produced by an algorithm that determines the largest set of parameters that has no significant difference between the parameters
- Factor mean ordering among groups and significant differences produced by z-tests

5.4 Alignment Example: Cross-Cultural Data on Nationalism and Patriotism

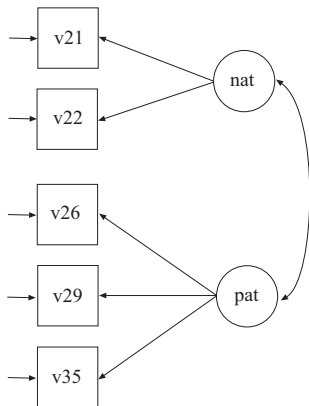
Davidov (2009). Measurement equivalence of nationalism and constructive patriotism in the ISSP: 34 countries in a comparative perspective. *Political Analysis*, 17, 64-82.

- Data from the International Social Survey Program (ISSP) 2003 National Identity Module
- 34 countries, n=45,546
- 5 measurements of nationalism and patriotism
- Expected 2-factor structure

Nationalism and Patriotism Data: Item Wording

- Nationalism factor:
 - V21: The world would be a better place if people from other countries were more like in [own country]
 - V22: Generally speaking, [own country] is better than most other countries
- Constructive Patriotism factor:
 - V26: How proud are you of [respondent's country] in the way democracy works?
 - V29: How proud are you of [respondent's country] in its social security system?
 - V35: How proud are you of [respondent's country] in its fair and equal treatment of all groups in society?

Nationalism and Patriotism Data: Confirmatory Factor Analysis (CFA) Model



	nat	pat
v21	x	0
v22	x	0
v26	0	x
v27	0	x
v35	0	x

Nationalism and Patriotism Data: Multiple-Group CFA with ML ($n = 45,546$)

Two-factor CFA with scalar measurement invariance across all 34 countries: $\chi^2(334) = 9669$, $p = 0$, RMSEA = 0.144, CFI = 0.721

Group-specific misfit evenly spread over the countries

Modification indices show a multitude of similarly large values

The usual multiple-group CFA approach fails

5.5 Input for Nationalism & Patriotism Alignment in 34 Countries

```
DATA:          FILE = issp.txt;
VARIABLE:     NAMES = country v21 v22 v26 v29 v35;
              USEVARIABLES = v21-v35;
              MISSING = v21-v35 (0 8 9);
              CLASSES = c(34);
              !KNOWNCLASS = c(country = 1 2 4 6-8 10-22 24-28 30-33 36
              !40-43);
              KNOWNCLASS = c(country);
ANALYSIS:     TYPE = MIXTURE;
              ESTIMATOR = ML;
              ALIGNMENT = FREE;
MODEL:        %OVERALL%
              nat BY v21-v22;
              pat BY v26-v35;
OUTPUT:       TECH1 TECH8 ALIGN;
```

- STANDARD ERROR COMPARISON INDICATES THAT THE FREE ALIGNMENT MODEL MAY BE POORLY IDENTIFIED. USING THE FIXED ALIGNMENT OPTION MAY RESOLVE THIS PROBLEM.

Choosing group with smallest factor mean to be the reference groups, this leads to the fixed alignment run:

ANALYSIS:

TYPE = MIXTURE;
ESTIMATOR = ML;
ALIGNMENT = FIXED(28);

Nationalism and Patriotism Example: Alignment Results

Approximate Measurement (**Non-**) Invariance by Group

Intercepts for Nationalism indicators (V21, V22) and Patriotism indicators (V26, V29, V35)

V21	1	2	3	4	5	6	7	8	9	10	11	12
	13	14	15	16	17	18	19	20	21	22	23	24
	25	26	27	28	29	30	31	32	33	34		
V22	(1)	2	3	(4)	5	(6)	7	8	(9)	10	11	12
	13	14	(15)	(16)	17	18	(19)	(20)	21	(22)	(23)	24
	(25)	26	27	28	(29)	30	31	(32)	33	34		
V26	1	2	3	4	5	6	7	8	9	10	11	12
	13	14	15	16	17	18	19	20	21	22	23	24
	25	26	27	28	29	30	31	32	33	34		
V29	(1)	2	3	(4)	(5)	6	7	(8)	(9)	10	11	12
	(13)	14	15	16	(17)	18	(19)	(20)	(21)	(22)	(23)	(24)
	(25)	26	27	28	29	(30)	31	32	33	(34)		
V35	(1)	(2)	3	(4)	5	6	7	(8)	(9)	(10)	11	12
	13	14	15	16	17	18	(19)	(20)	21	(22)	23	(24)
	25	26	(27)	(28)	(29)	(30)	31	32	(33)	34		

Nationalism and Patriotism Example: Alignment Results

Loadings for NATIONALISM factor

V21	1	(2)	(3)	4	5	6	7	(8)	(9)	(10)	11	12
	13	14	15	16	17	18	19	20	21	22	(23)	(24)
	(25)	26	27	28	29	(30)	31	32	33	34		
V22	1	2	3	4	5	6	7	8	9	10	11	12
	13	14	15	16	17	18	19	20	21	22	23	24
	25	26	27	28	29	30	31	32	33	34		

Loadings for PATRIOTISM factor

V26	1	2	3	4	5	6	7	8	9	10	11	12
	13	14	15	16	17	18	19	20	(21)	(22)	23	24
	25	26	27	(28)	29	30	31	32	33	34		
V29	1	2	3	4	5	6	7	8	9	10	11	12
	13	14	15	16	17	18	(19)	20	21	22	23	(24)
	25	26	27	28	29	30	31	32	33	34		
V35	1	2	3	4	5	6	7	8	9	10	11	12
	13	14	15	16	17	18	19	20	21	22	23	24
	25	26	27	28	29	30	31	32	33	34		

Nationalism and Patriotism Example: Factor Mean Comparisons (5% Significance Level)

Results for NATIONALISM factor

Ranking	Group	Value	Groups with significantly smaller factor mean
1	22	0.067	2 19 11 12 9 24 23 10 15 20 33 14 32 29 13 7 6 8 16 4 21 1 26 27 34 30 31 3 25 5
2	28	0.000	19 11 12 9 24 23 15 20 33 14 32 29 13 7 6 8 16 4 21 1 26 27 34 30 31 3 25 5 18 17
3	2	-0.284	6 16 4 21 1 26 27 34 31 3 25 5 18 17
4	19	-0.333	32 13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17
5	11	-0.344	33 32 13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17
6	12	-0.352	13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17
7	9	-0.357	7 6 16 4 21 1 26 27 34 31 3 25 5 18 17
8	24	-0.379	6 16 4 21 1 26 27 34 31 3 25 5 18 17
9	23	-0.388	13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17
10	10	-0.395	16 4 21 1 26 27 34 31 3 25 5 18 17
11	15	-0.396	13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17
12	20	-0.413	13 7 6 16 4 21 1 26 27 34 31 3 25 5 18 17

5.6 Alignment Monte Carlo Studies: How Do We Know That We Can Trust The Alignment Results?

- Simulations in Asparouhov-Muthén Web Note 18 and Muthén-Asparouhov (2014)
- Simulations based on the estimated model:
 - Request SVALUES for real-data alignment run (parameter estimates arranged as starting values)
 - Do a Monte Carlo run with these parameter values as population values, choosing the sample size and check parameter bias, SE bias, and the coverage
 - Do a "real-data" run on Monte-Carlo generated data from one or more replications to study the measurement invariance assessment - does it look like the real-data run?

Input for Alignment Monte Carlo Study

- Copy SVALUES results from real-data run into Monte Carlo run
- Delete [g#] statements, do a global change of the class label "c#" to "g#", and change f BY in OVERALL to give starting values

```
MONTECARLO:  NAMES = ipfrule ipmodst ipbhprp imptrad;
              NGROUPS= 26;
              NOBSERVATIONS = 26(2000);
              NREPS = 100;
              REPSAVE = ALL;
              SAVE = n2000f-22rep*.dat;
ANALYSIS:    TYPE = MIXTURE;
              ESTIMATOR = ML;
              ALIGNMENT = FIXED(22);
              PROCESSORS = 8;
MODEL POPULATION:
              %OVERALL%
              traco BY ipfrule-imptrad*1;
              ! [ g#1*-0.10053 ];
              etc
```

- Monte Carlo output gives the average correlation over replications for the group-specific population factor means and estimated factor means
- A correlation > 0.98 reflects a very similar ordering of the groups with respect to factor means
- Correlations can be high even with a high degree of measurement non-invariance

5.7 Alignment Optimization: Binary Math Items in 40 Countries (PISA)

- Items from the PISA (Program for International Student Assessment) survey of 2003
- A total of 9796 students from 40 countries
- Analyzed by Fox (2010). Bayesian Item Response Modeling
- A 40-group, one-factor model for eight mathematics test items
- 2-parameter probit IRT model that accommodates country measurement non-invariance for all difficulty (threshold) and discrimination (loading) parameters as well as country-specific factor means and variances

Input for PISA Alignment with Binary Items using Bayes

```
DATA:          FILE = pisa2003.dat;
VARIABLE:      NAMES = cn y1-y8;
                CATEGORICAL = y1-y8;
                USEVARIABLES = y1-y8;
                MISSING = y1-y8(9);
                CLASSES = c(40);
                KNOWNCLASS = c(cn = 1-40);
ANALYSIS:      TYPE = MIXTURE;
                ESTIMATOR = BAYES;
                PROCESSORS = 2;
                ALIGNMENT = FREE;
                THIN = 10; ! record only every 10th iter; saves alignment time
                BITERATIONS = (5000); ! do a minimum of 5000 iterations
MODEL:         %OVERALL%
                f BY y1-y8;
OUTPUT:        TECH1 TECH8 ALIGN;
PLOT:          TYPE = PLOT2;
```

5.8 Multiple-Group Analysis using Bayes and BSEM Alignment

The several uses of BSEM with zero-mean, small-variance priors:

- Single group analysis (2012 Psych Methods article):
 - Cross-loadings
 - Residual covariances
 - Direct effects in MIMIC
- Multiple-group analysis:
 - Configural and scalar analysis with cross-loadings and/or residual covariances
 - Approximate measurement invariance (Web Note 17)
 - BSEM-based alignment optimization (Web Note 18):
 - Residual covariances
 - Approximate measurement invariance

What does Bayes contribute?

- 1 Bayes with informative, zero-mean, small-variance priors for residual covariances can allow better configural fit - configural misfit in some groups is a common problem
- 2 Bayes with informative, zero-mean, small-variance priors for measurement parameter differences across groups (multiple-group BSEM) can allow better scalar fit
 - MG-BSEM as an alternative to alignment (finds non-invariance)
 - MG-BSEM-based alignment (advantageous for small samples?)
- 3 Bayes alignment can produce plausible values for the subjects' factor score values to be used in further analyses

- ML estimation:
 - ALIGNMENT = FREE
 - ALIGNMENT = FIXED(value)
- Bayes estimation:
 - ALIGNMENT = FREE
 - ALIGNMENT = FIXED(group)
 - ALIGNMENT = FREE(BSEM) - "BSEM-based alignment"
 - ALIGNMENT = FIXED(group BSEM)
 - Adding Inverse Wishart (IW) priors for Theta to allow residual covariances

Muthén & Asparouhov (2013). BSEM measurement invariance analysis. Web Note 17.

- Approximate measurement invariance across groups using zero-mean, small-variance informative priors for the group differences
- Produces "modification indices" by flagging non-invariant items as significantly deviating from average (ML-based MIs not available for categorical items)
- Freeing the non-invariant parameters gives proper "alignment", otherwise an alignment run is needed (BSEM-based alignment: ALIGNMENT = FREE(BSEM);)

Multiple-Group BSEM: Math Items in 40 PISA Countries

```
DATA:          FILE = pisa2003.dat;
VARIABLE:      NAMES = cn y1-y8;
                CATEGORICAL = y1-y8;
                USEVARIABLES = y1-y8;
                MISSING = y1-y8(9);
                CLASSES = c(40);
                KNOWNCLASS = c(cn = 1-40);
ANALYSIS:      TYPE = MIXTURE;
                ESTIMATOR = BAYES;
                PROCESSORS = 2;
                MODEL = ALLFREE ;
                ITERATIONS = (10000);
MODEL:         %OVERALL%
                f BY y1-y8* (lam#_1-lam#_8);
                [y1$1-y8$1] (tau#_1-tau#_8);
                %c#40%
                [f@0];
                f@1;
```

Multiple-Group BSEM: Math Items in 40 PISA Countries, Continued

MODEL PRIORS:

DO(1,8) DIFF(tau1_#-tau40_#)~N(0,0.10);

DO(1,8) DIFF(lam1_#-lam40_#)~N(0,0.10);

OUTPUT: TECH1 TECH2;

PLOT: TYPE = PLOT2;

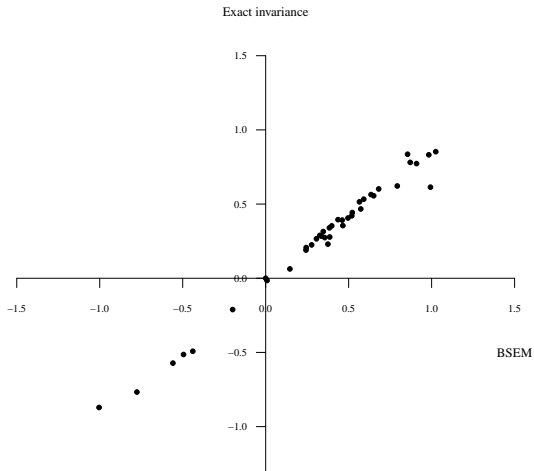
Multiple-Group BSEM: Non-Invariance Findings for PISA Items

Table : PISA countries with significant differences relative to the average across countries (prior variance = 0.10)

Item	Loading	Threshold
1	-	2, 12, 18, 22, 28, 39
2	15, 35	29, 38
3	15	23, 34, 35
4	-	12, 27, 40
5	3	7, 37
6	3, 33	5, 18, 25, 27, 37
7	-	9, 24, 27
8	24	-

Estimated Factor Means for 40 PISA Countries

Figure : Estimated factor means for 40 countries: Comparing BSEM analysis (X axis) with analysis imposing exact invariance (Y axis)



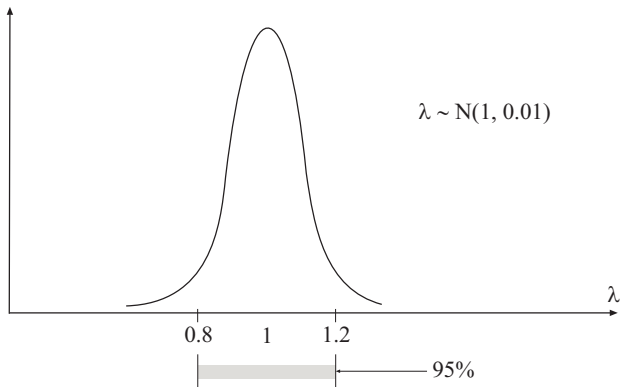
Groups seen as random clusters

- De Jong, Steenkamp & Fox (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. *Journal of Consumer Research*, 34, 260-278.
- Fox (2010). *Bayesian Item Response Modeling*. Springer
- Fox & 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*
- Asparouhov & Muthén (2012). General random effect latent variable modeling: Random subjects, items, contexts, and parameters
- Bayesian estimation needed because random loadings with ML give rise to numerical integration with many dimensions

Two-Level Analysis with Random Item Parameters: A New Conceptualization of Measurement Invariance

Each measurement parameter varies across groups/clusters, but groups/clusters have a common mean and variance. E.g.

$$\lambda_j \sim N(\mu_\lambda, \sigma_\lambda^2). \quad (30)$$



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

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

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

This is a 2-parameter probit IRT model where both discrimination (a) and difficulty (b) vary across country

- The θ ability factor is decomposed as

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

- The mean and variance of the ability vary across country
- Model preserves common measurement scale while accommodating measurement non-invariance
- The ability for each country obtained by factor score estimation

5.11 Random Loadings: UG Ex9.19

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

```
TITLE:      this is an example of a two-level MIMIC
            model with continuous factor indicators,
            random factor loadings, two covariates on
            within, and one covariate on between
            with equal loadings across levels

DATA:      FILE = ex9.19.dat;

VARIABLE:  NAMES = y1-y4 x1 x2 w clus;
            WITHIN = x1 x2;
            BETWEEN = w;
            CLUSTER = clus;

ANALYSIS:  TYPE = TWOLEVEL RANDOM;
            ESTIMATOR = BAYES;
            PROCESSORS = 2;
            BITER = (1000);

MODEL:     %WITHIN%
            s1-s4 | f BY y1-y4;
            f@1;
            f ON x1 x2;
            %BETWEEN%
            f ON w;
            f; ! defaults: s1-s4; [s1-s4];

PLOT:      TYPE = PLOT2;

OUTPUT:    TECH1 TECH8;
```

5.12 Alignment Summary

- Multiple groups/clusters data can be represented by fixed or random mode models
 - Having many groups/clusters does not preclude fixed-mode, multiple-group analysis
- Fixed mode modeling can explore the data using non-identified models:
 - Alignment optimization methods (Asparouhov-Muthén, Web Note 18)
 - Bayesian (BSEM) methods (Muthén-Asparouhov, Web Note 17)
- Random mode, two-level modeling:
 - Conventional two-level factor analysis reveals some limited forms of non-invariance (intercepts)
 - Random slope two-level factor analysis reveals more general forms of non-invariance

- Fixed mode modeling using alignment optimization has many advantages over random mode modeling:
 - Convenient, one-step analysis
 - Points to which groups/clusters contribute to non-invariance
 - Is not limited to just > 30 clusters, but works well with any number of groups/clusters (say < 100 , or say $< 3,000$ configural parameters)
 - Gives an ordering of the factor means without having to estimate factor scores for each group/cluster
 - Allows factor variance variation across groups/clusters without involving random slopes
 - Does not assume normally-distributed non-invariance

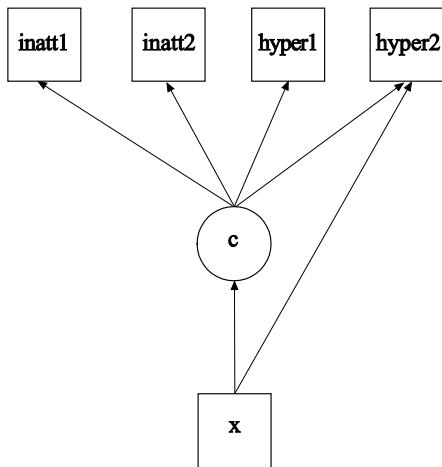
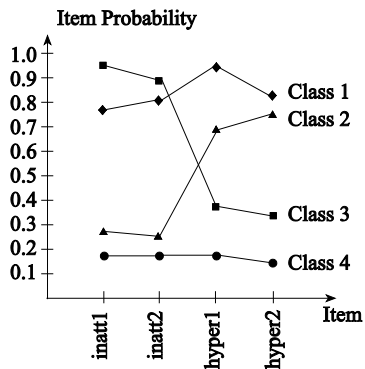
- The big news: Alignment optimization:
 - Does modeling with group-specific measurement intercepts, measurement loadings, factor means, and factor variances
 - Aligns to minimal measurement non-invariance
 - Uses EFA-like tools to identify non-identified parameters
 - Is easy to do
- The other news: The Alignment optimization companion technique - multiple-group BSEM
- All available in Mplus starting with Version 7.11

Analysis Methods

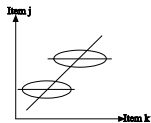
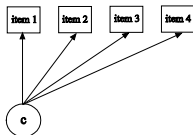
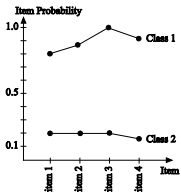
- Regression mixture models - Modeling of counts, randomized interventions with non-compliance (CACE)
- Latent class analysis with and without covariates
- Latent transition analysis
- Latent class growth analysis
- Growth mixture modeling
- Survival mixture modeling

- 3-step mixture modeling: Analyze-classify-analyze approaches to investigate covariates and distal outcomes
 - LCA
 - Regression mixture analysis
 - GMM
 - LTA
- Latent transition analysis (LTA)
 - Introductory examples
 - New Mplus output
 - Covariates influencing transition probabilities
 - Probability parameterization useful for Mover-Stayer LTA
 - LTA extensions
- Residual association parameters (version 7.2)
- Non-normal within-class distributions (version 7.2)

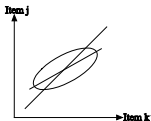
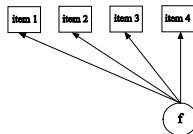
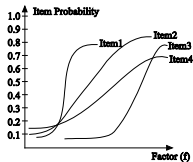
6.1 Latent Class Analysis



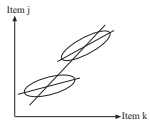
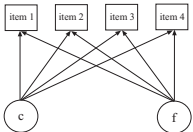
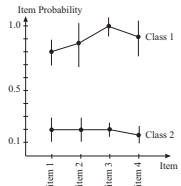
Latent Class Analysis



Factor Analysis (IRT)



Factor Mixture Analysis



Latent Class, Factor, And Factor Mixture Analysis Alcohol Dependence Criteria, NLSY 1989 (n = 8313)

Source: Muthén & Muthén (1995)

	Latent Classes				
	Two-class solution ¹		Three-class solution ²		
	I	II	I	II	III
Prevalence	0.78	0.22	0.75	0.21	0.03
DSM-III-R criterion	conditional probability of fulfilling a criterion				
Withdrawal	0.00	0.14	0.00	0.07	0.49
Tolerance	0.01	0.45	0.01	0.35	0.81
Larger	0.15	0.96	0.12	0.94	0.99
Cut down	0.00	0.14	0.01	0.05	0.60
Time spent	0.00	0.19	0.00	0.09	0.65
Major role-hazard	0.03	0.83	0.02	0.73	0.96
Give up	0.00	0.10	0.00	0.03	0.43
Relief	0.00	0.08	0.00	0.02	0.40
Continue	0.00	0.24	0.02	0.11	0.83

¹Likelihood ratio chi-square fit = 1779 with 492 degrees of freedom

²Likelihood ratio chi-square fit = 448 with 482 degrees of freedom

- LCA, 3 classes: $\log L = -14,139$, 29 parameters, BIC = 28,539
- FA, 2 factors: $\log L = -14,083$, 26 parameters, BIC = 28,401
- FMA 2 classes, 1 factor, loadings invariant:
 $\log L = -14,054$, 29 parameters, BIC = 28,370

Models can be compared with respect to fit to the data:

- Standardized bivariate residuals
- Standardized residuals for most frequent response patterns

Estimated Frequencies And Standardized Residuals

Obs. Freq.	LCA 3c		FA 2f		FMA 1f, 2c	
	Est. Freq.	Res.	Est. Freq.	Res.	Est. Freq.	Res.
5335	5332	-0.07	5307	-0.64	5331	-0.08
941	945	0.12	985	1.48	946	0.18
601	551	-2.22	596	-0.22	606	0.21
217	284	4.04	211	-0.42	228	0.75
155	111	-4.16	118	-3.48	134	1.87
149	151	0.15	168	1.45	147	0.17
65	68	0.41	46	-2.79	53	1.60
49	52	0.42	84	3.80	59	1.27
48	54	0.81	44	-0.61	46	0.32
47	40	-1.09	45	-0.37	45	0.33

Bolded entries are significant at the 5% level.

TITLE: Alcohol LCA M & M (1995)
DATA: FILE = bengt05_spread.dat;
VARIABLE: NAMES = u1-u9;
CATEGORICAL = u1-u9;
CLASSES = c(2);
ANALYSIS: TYPE = MIXTURE;
ALGORITHM = INTEGRATION;
STARTS = 200 10; STITER = 20;
ADAPTIVE = OFF;
PROCESSORS = 4(STARTS);

Input For FMA Of 9 Alcohol Items In The NLSY 1989 (Continued)

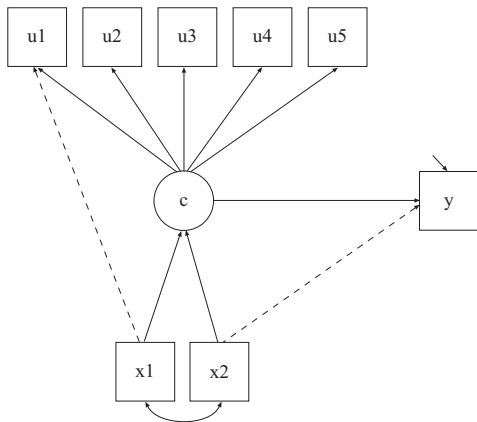
MODEL: %OVERALL%
 f BY u1-u9;
 f*1; [f@0];
 %c#1%
 [u1\$1-u9\$1];
 f*1;
 %c#2%
 [u1\$1-u9\$1];
 f*1;

OUTPUT: TECH1 TECH8 TECH10;

PLOT: TYPE = PLOT3;
 SERIES = u1-u9(*);

6.2 3-Step Mixture Modeling

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



1-Step vs 3-Step: A Hypothetical Genetic Example

Substantive question: Should the latent classes be defined by the indicators alone or also by covariates and distal outcomes (antecedents and consequences)?

- Example: Study of genotypes (x variables) influencing phenotypes (y variables)
- 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

- Latent class models should be subjected to both statistical and substantive checking (Muthén, 2003 in Psychological Methods)
- Substantive checking can be done by relating latent classes to antecedents and consequences (covariates and distal outcomes)
- The 3-step approach is a useful tool for this

The Old 3-Step Approach

- 1 Estimate the LCA model
- 2 Determine each subject's most likely class membership
- 3 Relate the most likely class variable to other variables

The old 3-step approach is problematic: Unless the classification is very good (high entropy), this gives biased estimates and biased standard errors for the relationships with other variables.

The LCA Provides Information About the Classification Quality

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

	1	2	3
1	0.839	0.066	0.095
2	0.053	0.845	0.102
3	0.125	0.107	0.768

The New 3-Step Approach

- New Method in Mplus Version 7: 3-Step approach correcting for classification error
 - 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 misclassification rate of N estimated in the step 1 LCA analysis
- Bolck, Croon, & Hagenaars (2004) Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, 12, 3-27.
- Vermunt (2010). Latent Class Modeling with Covariates: Two improved three-step approaches. *Political Analysis*, 18, 450-469
- Asparouhov & Muthén (2012). Auxiliary variables in mixture modeling: A 3-step approach using Mplus. Mplus Web Note 15. *Structural Equation Modeling*, 21, 329-341

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

	1	2	3
1	0.839	0.066	0.095
2	0.053	0.845	0.102
3	0.125	0.107	0.768

$$\log(0.839/0.095) = 2.178$$

$$\log(0.066/0.095) = -0.364$$

$$\log(0.053/0.102) = -0.654$$

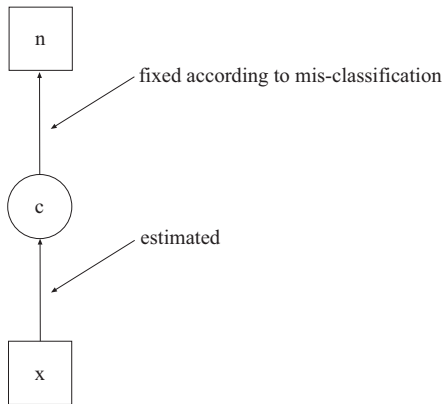
$$\log(0.845/0.102) = 2.114$$

$$\log(0.125/0.768) = -1.815$$

$$\log(0.107/0.768) = -1.970$$

Step 3 Regression on a Covariate

- n : Most likely class membership from Step 2 (nominal variable)
- c : Latent class variable
- x : Covariate



Input File for Step 3 in the 3-Step Estimation

```
VARIABLE:  NAMES = u1-u5 x p1-p3 n;  
           USEVARIABLES = x n;  
           CLASSES = c(3);  
           NOMINAL = n;  
DATA:      FILE = man3step2.dat;  
ANALYSIS:  TYPE = MIXTURE; STARTS = 0;  
MODEL:     %OVERALL%  
           c ON x;  
           %c#1%  
           [n#1@2.178];  
           [n#2@-0.364];  
           %c#2%  
           [n#1@-0.654];  
           [n#2@2.114];  
           %c#3%  
           [n#1@-1.815];
```



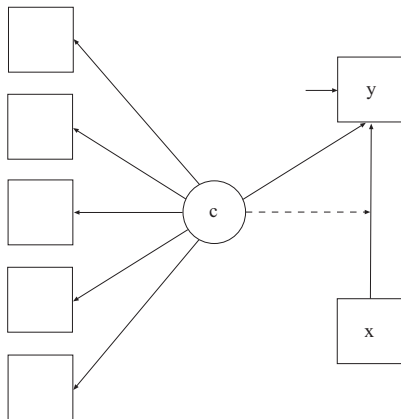
```
VARIABLE:  NAMES = u1-u5 x;  
           CATEGORICAL = u1-u5;  
           CLASSES = c(3);  
           AUXILIARY = x(R3STEP);  
DATA:      FILE = 3step.dat;  
ANALYSIS:  TYPE = MIXTURE;  
MODEL:     !no model is needed, LCA is default
```

```
VARIABLE:  NAMES = u1-u5 y;  
           CATEGORICAL = u1-u5;  
           CLASSES = c(3);  
           AUXILIARY = y(DU3STEP);  
DATA:      FILE = 3step.dat;  
ANALYSIS:  TYPE = MIXTURE;  
MODEL:     !no model is needed, LCA is default
```

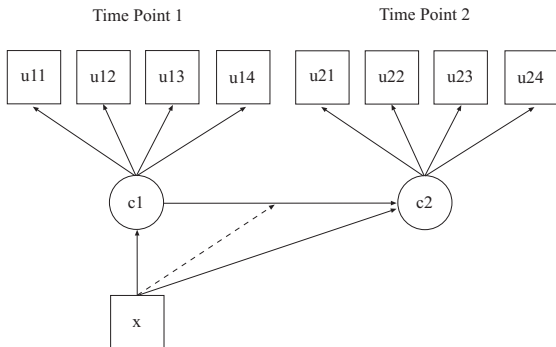
A Second Look at Distal 3-Step

- In some examples the Asparouhov-Muthén distal 3-step method in Mplus Web Note 15 leads to changes in latent class formation between Step 1 and Step 3 - warning given in Mplus Version 7.1
- Lanza et al. (2013) in the SEM journal propose a different distal 3-step method that avoids changes in class formation. Included in Mplus Version 7.1 (DCON/DCAT).
- Future research needed to evaluate which method, including Most Likely Class and Pseudo-class, is least sensitive to violations of assumptions such as no direct effects
- Bakk et al. (2014) forthcoming in SEM propose a method based on training data (best distal method so far?): BCH option in Mplus 7.21

Manual 3-Step Mixture Modeling For Special Models: A Regression Mixture Example

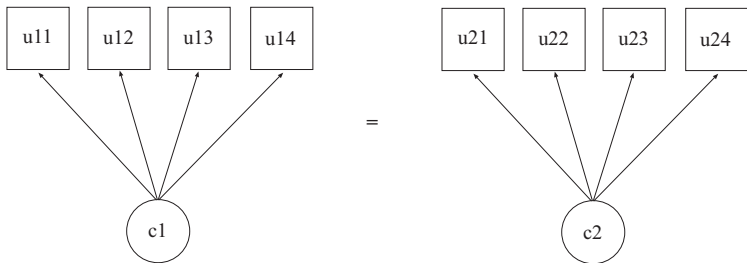


3-Step Latent Transition Analysis

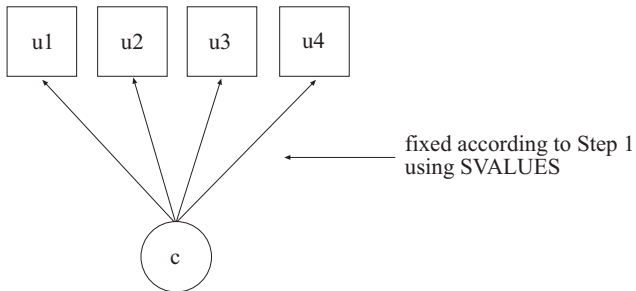


- Asparouhov-Muthén (2014): Auxiliary Variables in Mixture Modeling: 3-Step Approaches Using Mplus. Forthcoming in Structural Equation Modeling
- Nylund et al. (2014): A Latent Transition Mixture Model Using the Three-Step Specification. Forthcoming in Structural Equation Modeling

LTA: Step 1

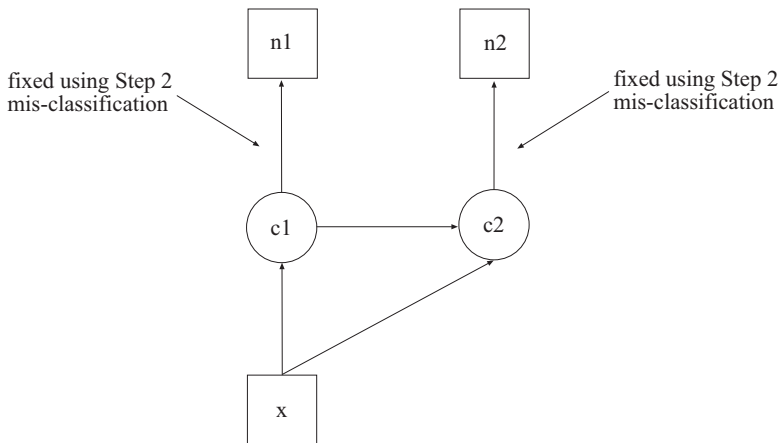


For each time point:

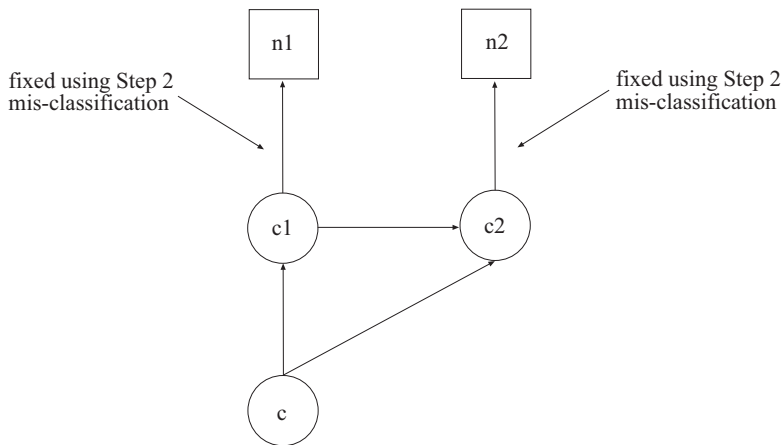


Save most likely class

LTA: Step 3



3-Step Mover-Stayer LTA



6.3 Latent Transition Analysis Developments

New developments in Version 7:

- TECH15 output with conditional class probabilities useful for studying transition probabilities varying as a function of an observed binary or nominal covariate such as treatment/control, ethnicity, 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
- New User's Guide examples
 - 8.13: LTA for two time points with a binary covariate influencing the latent transition probabilities
 - 8.14: LTA for two time points with a continuous covariate influencing the latent transition probabilities
 - 8.15: Mover-stayer LTA for three time points using a probability parameterization

6.4 Latent Class and Latent Transition Analysis with Residual Covariances (Residual Associations) for Categorical Items In Mplus Version 7.2

- Addition of a within-class two-way loglinear model, adding one association parameter per variable pair
 - Binary items: saturates the 2×2 table
 - Ordered polytomous items: Uniform association model (Goodman 1979)
- No need for numerical integration due to adding a factor behind the pair of items
- Association parameter can be equal of different across latent classes
- Covariates allowed, but not direct effects on items
- Asparouhov & Muthén (2014). Residual associations in latent class and latent transition analysis. Forthcoming in Structural Equation Modeling

LCA Example: Deciding On The Number Of Classes For 17 Antisocial Behavior Items ($n = 7326$)

Five-Class Solution

The five-class solution is substantively meaningful:

Class 1	138.06985	0.01888	High Overall
Class 2	860.41897	0.11771	Property Offense
Class 3	1257.56652	0.17151	Drugs
Class 4	1909.32749	0.26219	Person Offense
Class 5	3160.61717	0.42971	Normative (Pot)

Six-Class Solution - adds a variation on Class 2 in the 5-class solution

Deciding On The Number Of Classes For 17 ASB Items

Number of classes	1	2	3	4	5	6
Loglikelihood	-48168.475	-42625.653	-41713.142	-41007.498	-40808.312	-40604.231
# par.	17	35	53	71	89	107
BIC	96488	85563	83898	82647	82409	82161

TECH10 bivariate tests in the 5-class run show need for adding residual covariances. Adding 4 residual covariances to the 5-class model:

Loglikelihood = -40603, # parameters = 93, BIC = 82034

VARIABLE: NAMES = property fight shoplift lt50 gt50 force threat injure
pot drug soldpot solddrug con auto bldg goods gambling
dsm1-dsm22 sex black hisp single divorce dropout college
onset f1 f2 f3 age94 cohort dep abuse;
USEVARIABLES = property-gambling;
CATEGORICAL = property-gambling;
CLASSES = c(5);

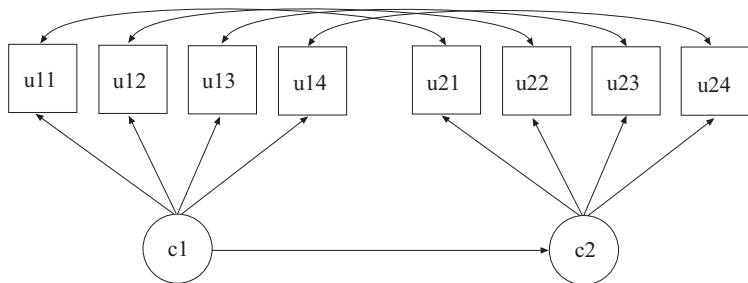
ANALYSIS: TYPE = MIXTURE;
STARTS = 1200 300;
PARAMETERIZATION = RESCOV;

MODEL: %OVERALL%
threat WITH injure;
drug WITH soldpot;
drug WITH solddrug;
soldpot WITH solddrug;

OUTPUT: TECH1 TECH8 TECH10;

Residual covariances can also be specified as class specific.

LTA with Correlated Residuals



Allowing across-time correlation for each item changes the estimated latent transition probabilities.

6.5 Overview of Mixture Modeling with Non-Normal Distributions in Mplus Version 7.2

- 1 A new growth mixture modeling (GMM) method
 - Examples of skew distributions
 - Normal mixtures
 - Introducing mixtures of non-normal distributions
 - Non-normal mixtures of latent variable models:
 - GMM of BMI in the NLSY multiple-cohort study
 - Math and high school dropout in the LSAY study
 - Cat's cradle concern
 - Disadvantages and advantages of non-normal mixtures
 - Mplus specifications
- 2 A new SEM method: Non-normal SEM
 - Path analysis
 - Factor analysis
 - SEM

Mixture Modeling with Non-Normal Distributions in Mplus Version 7.2

References:

- Asparouhov & Muthén (2014). Structural equation models and mixture models with continuous non-normal skewed distributions. Forthcoming in Structural Equation Modeling
- Muthén & Asparouhov (2014). Growth mixture modeling with non-normal distributions. Submitted

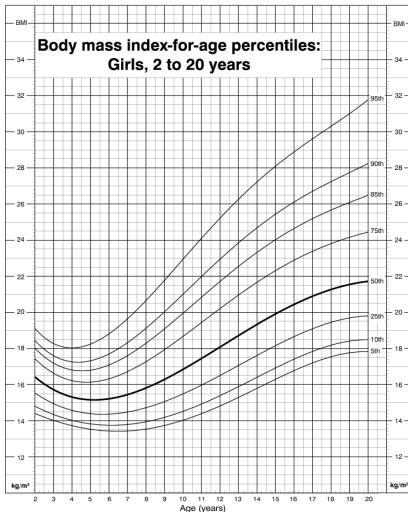
6.6 Examples of Skewed Distributions

- Body Mass Index (BMI) in obesity studies (long right tail)
- Mini Mental State Examination (MMSE) cognitive test in Alzheimer's studies (long left tail)
- PSA scores in prostate cancer studies (long right tail)
- CD4 cell counts in AIDS studies (long right tail)
- Ham-D score in antidepressant studies (long right tail)

Body Mass Index (BMI): kg/m^2

Normal $18 < BMI < 25$, Overweight $25 < BMI < 30$, Obese > 30

CDC Growth Charts: United States



Published May 30, 2000.
SOURCE: Developed by the National Center for Health Statistics in collaboration with
the National Center for Chronic Disease Prevention and Health Promotion (2000).



SAFER • HEALTHIER • PEOPLE™

NLSY Multiple-Cohort Data Ages 12 to 23

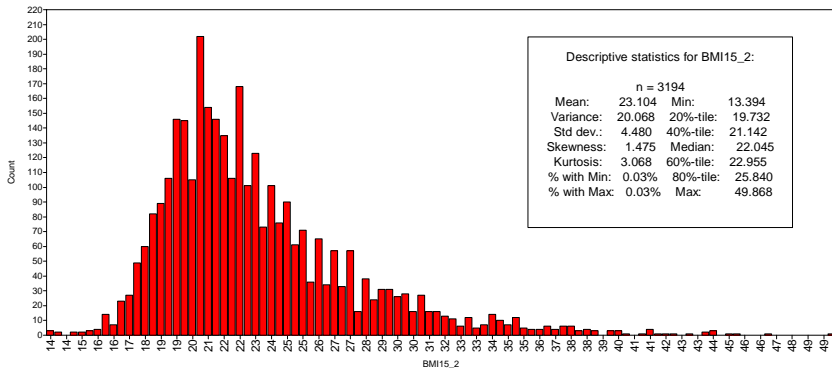
Accelerated longitudinal design - NLSY97

	12	13	14	15	16	17	18	19	20	21	22	23
1997	1,165	1,715	1,847	1,868	1,709	613						
1998		104	1,592	1,671	1,727	1,739	1,400	106				
1999			108	1,659	1,625	1,721	1,614	1,370	65			
2000				57	1,553	1,656	1,649	1,597	1,390	132		
2001					66	1,543	1,615	1,602	1,582	1,324	109	
2002							1,614	1,587	1,643	1,582	1,324	106
2003							112	1,497	1,600	1,582	1,564	1,283
Totals	1,165	1,819	3,547	5,255	6,680	7,272	8,004	7,759	6,280	4,620	2,997	1,389

NLSY, National Longitudinal Survey of Youth

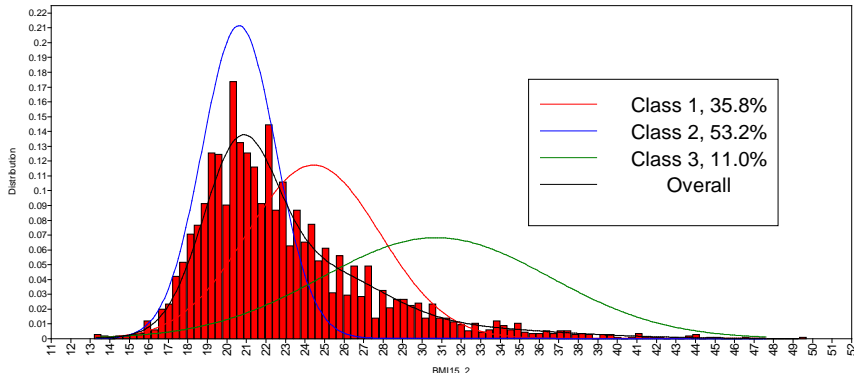
Source: Nonnemaker et al. (2009). Youth BMI trajectories: Evidence from the NLSY97, Obesity

BMI at Age 15 in the NLSY (Males, $n = 3194$)



6.7 Mixtures for Male BMI at Age 15 in the NLSY

- Skewness = 1.5, kurtosis = 3.1
- Mixtures of normals with 1-4 classes have BIC = 18,658, 17,697, 17,638, 17,637 (tiny class)
- 3-class mixture shown below

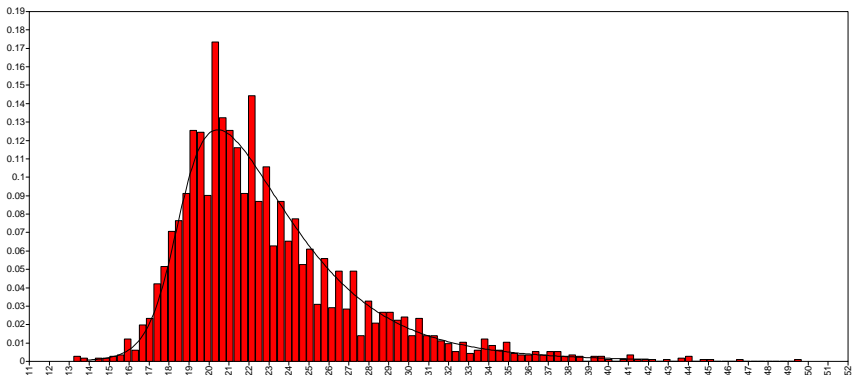


Several Classes or One Non-Normal Distribution?

- Pearson (1895)
- Hypertension debate:
 - Platt (1963): Hypertension is a "disease" (separate class)
 - Pickering (1968): Hypertension is merely the upper tail of a skewed distribution
- Schork & Schork (1988): Two-component mixture versus lognormal
- Bauer & Curran (2003): Growth mixture modeling classes may merely reflect a non-normal distribution so that classes have no substantive meaning
- Muthén (2003) comment on BC: Substantive checking of classes related to antecedents, concurrent events, consequences (distal outcomes), and usefulness
- Multivariate case more informative than univariate

What If We Could Instead Fit The Data With a Skewed Distribution?

- Then a mixture would not be necessitated by a non-normal distribution, but a single class may be sufficient
- A mixture of non-normal distributions is possible



6.8 Introducing Mixtures of Non-Normal Distributions in Mplus Version 7.2

In addition to a mixture of normal distributions, it is now possible to use

- T: Adding a degree of freedom (df) parameter (thicker tails than normal; $df > 30 \approx$ normal)
- Skew-normal: Adding a skew parameter to each variable
- Skew-T: Adding skew and df parameters (stronger skew possible than skew-normal)

References

- Azzalini (1985), Azzalini & Dalla Valle (1996): skew-normal
- Arellano-Valle & Genton (2010): extended skew-t
- McNicholas, Murray, 2013, 2014: skew-t as a special case of the generalized hyperbolic distribution
- McLachlan, Lee, Lin, 2013, 2014: restricted and unrestricted skew-t

Skew T-Distribution Formulas

Y can be seen as the sum of a mean, a part that produces skewness, and a part that adds a symmetric distribution:

$$Y = \mu + \delta|U_0| + U_1,$$

where U_0 has a univariate t and U_1 a multivariate t distribution. Expectation, variance (δ is a skew vector, ν the df):

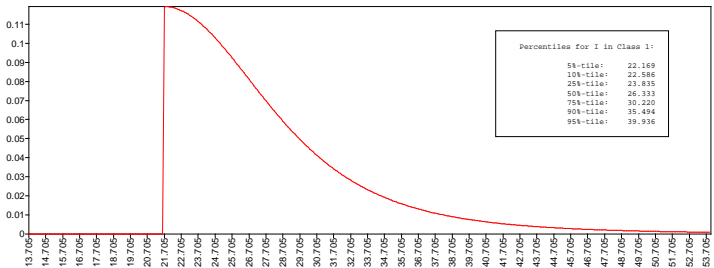
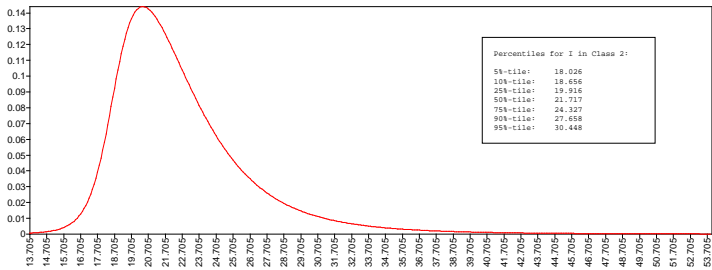
$$E(Y) = \mu + \delta \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \sqrt{\frac{\nu}{\pi}},$$

$$\text{Var}(Y) = \frac{\nu}{\nu-2} (\Sigma + \delta\delta^T) - \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^2 \frac{\nu}{\pi} \delta\delta^T$$

Marginal and conditional distributions:

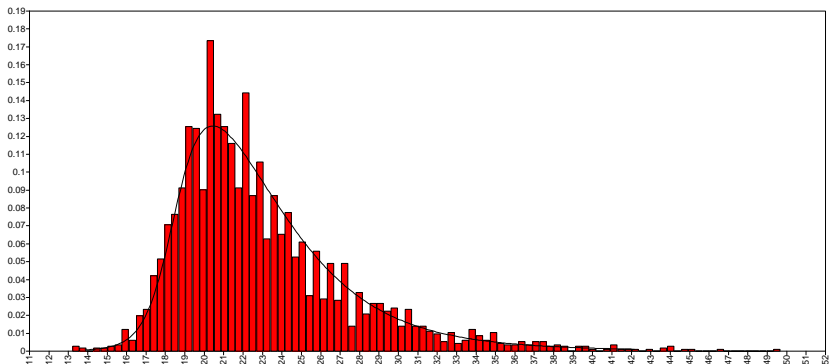
- Marginal is also a skew-t distribution
- Conditional is an extended skew-t distribution

Examples of Skew-T Distributions



BMI at Age 15 in the NLSY (Males, $n = 3194$)

- Skewness = 1.5, kurtosis = 3.1
- Mixtures of normals with 1-4 classes: BIC = 18,658, 17,697, 17,638, 17,637 (tiny class). 3-class model uses 8 parameters
- 1-class Skew-T distribution: BIC = 17,623 (2-class BIC = 17,638). 1-class model uses 4 parameters



6.9 Cluster Analysis by "Mixtures of Factor Analyzers" (McLachlan)

Reduces the number of μ_c, Σ_c parameters for $c = 1, 2, \dots, C$ by applying the Σ_c structure of an EFA with orthogonal factors:

$$\Sigma_c = \Lambda_c \Lambda_c' + \Theta_c \quad (31)$$

This leads to 8 variations by letting Λ_c and Θ_c be invariant or not across classes and letting Θ_c have equality across variables or not (McNicholas & Murphy, 2008).

Interest in clustering as opposed to the factors, e.g. for genetic applications.

(EFA mixtures not yet available in Mplus for non-normal distributions, but can be done using "EFA-in-CFA".)

Models:

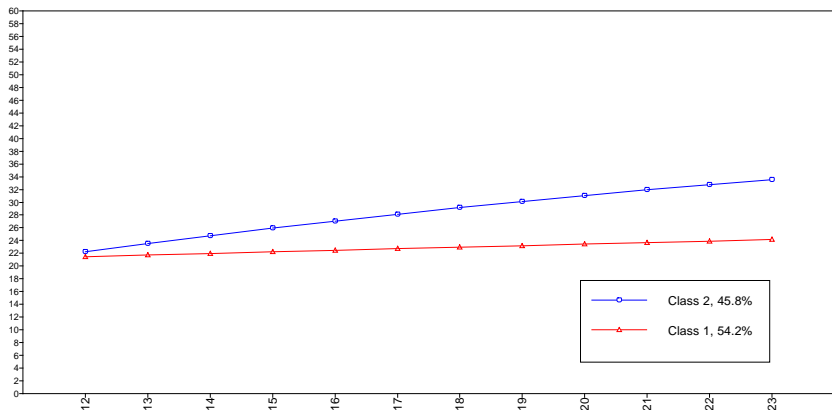
- Mixtures of Exploratory Factor Models (McLachlan, Lee, Lin; McNicholas, Murray)
- Mixtures of Confirmatory Factor Models; FMM (Mplus)
- Mixtures of SEM (Mplus)
- Mixtures of Growth Models; GMM (Mplus)

Choices:

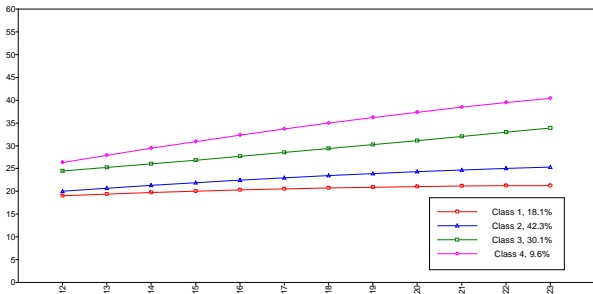
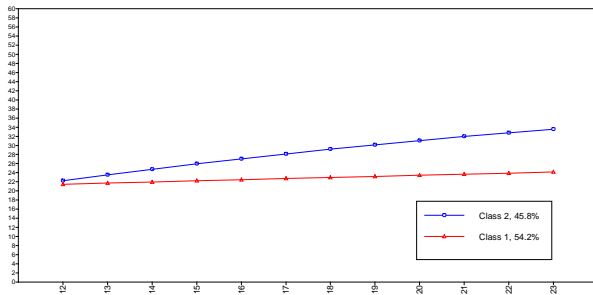
- Intercepts, slopes (loadings), and residual variances invariant?
- Scalar invariance (intercepts, loadings) allows factor means to vary across classes instead of intercepts (not typically used in mixtures of EFA, but needed for GMM)
- Skew for the observed or latent variables? Implications for the observed means. Latent skew suitable for GMM - the observed variable means are governed by the growth factor means

6.11 Growth Mixture Modeling of NLSY BMI Age 12 to 23 for Black Females ($n = 1160$)

- Normal BIC: 31684 (2c), 31386 (3c), **31314 (4c)**, 31338 (5c)
- Skew-T BIC: 31411 (1c), **31225 (2c)**, 31270 (3c)

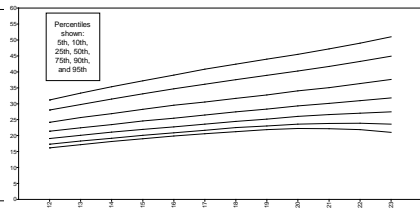
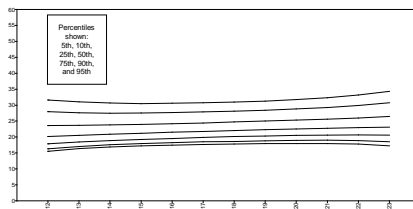


2-Class Skew-T versus 4-Class Normal

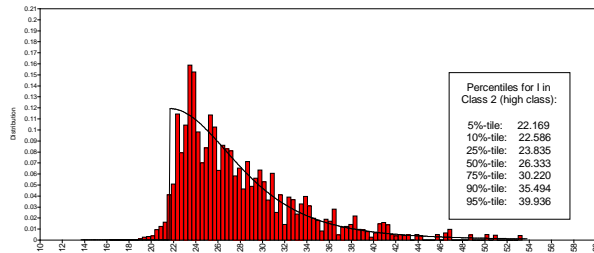
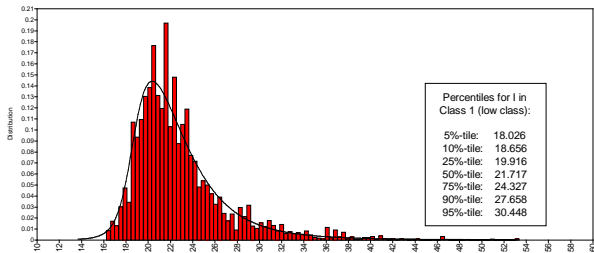


2-Class Skew-T: Estimated Percentiles

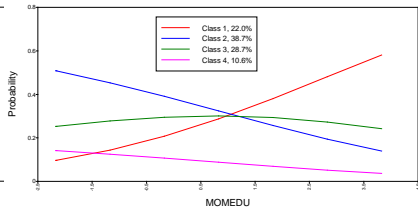
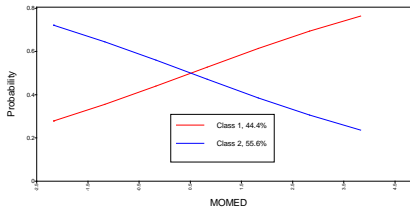
(Note: Not Growth Curves)



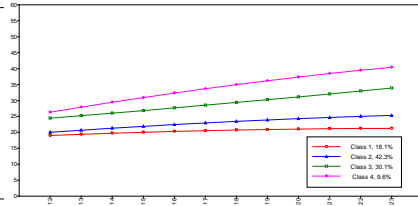
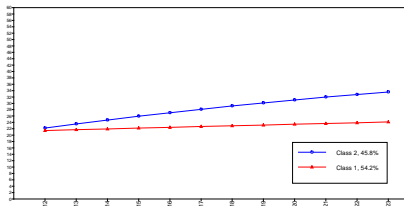
2-Class Skew-T: Intercept Growth Factor (Age 17)



Regressing Class on a MOMED Covariate ("c ON x"): 2-Class Skew-T versus 4-Class Normal

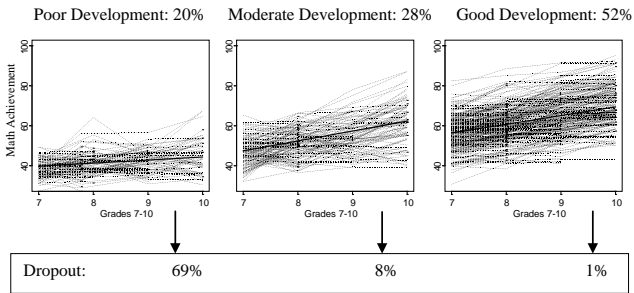


Recall the estimated trajectory means for skew-t versus normal:



6.12 Growth Mixture Modeling: Math Achievement Trajectory Classes and High School Dropout.

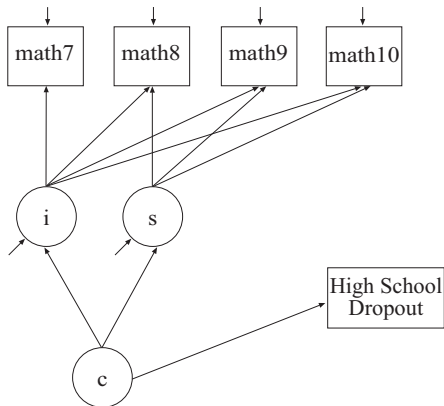
An Example of Substantive Checking via Predictive Validity



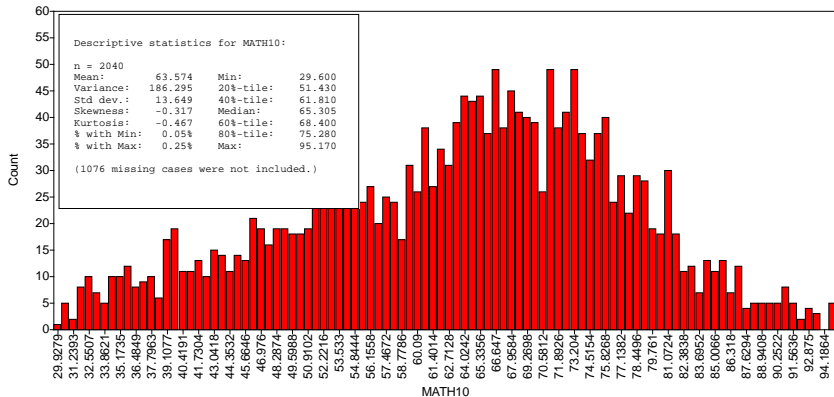
Source: Muthén (2003). Statistical and substantive checking in growth mixture modeling. *Psychological Methods*.

- Does the normal mixture solution hold up when checking with non-normal mixtures?

Growth Mixture Modeling: Math Achievement Trajectory Classes and High School Dropout



Growth Mixture Modeling: Math Achievement Trajectory Classes and High School Dropout



Growth Mixture Modeling: Math Achievement Trajectory Classes and High School Dropout

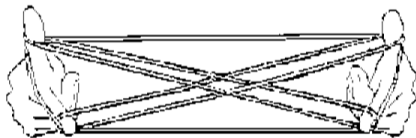
Best solutions, 3 classes (LL, no. par's, BIC):

- Normal distribution: -34459, 32, **69175**
- T distribution: -34453, 35, 69188
- Skew-normal distribution: -34442, 38, 69191
- Skew-t distribution: -34439, 42, 69207

Percent in low, flat class and odds ratios for dropout vs not, comparing low, flat class with the best class:

- Normal distribution: 18 %, OR = 17.1
- T distribution: 19 %, OR = 20.6
- Skew-normal distribution: 26 %, OR = 23.8
- Skew-t distribution: 26 %, OR = 37.3

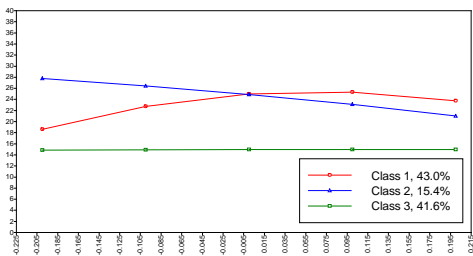
6.13 Cat's Cradle Concern



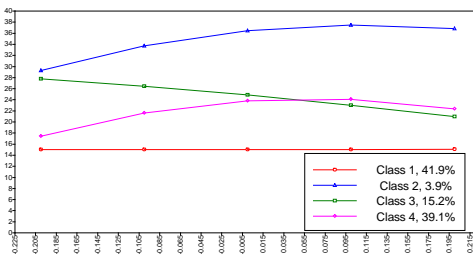
Source: Sher, Jackson, Steinley (2011). Alcohol use trajectories and the ubiquitous cat's cradle: Cause for concern? *Journal of Abnormal Psychology*.

Cat's Cradle Concern: Generated Data ($n = 2,000$)

Data from a 3-class skew-t (S/K=1.5/4). 3-class skew-t, BIC=43566:



4-class normal - cat's cradle with a high/chronic class, BIC=44935:



6.14 Disadvantages of Non-Normal Mixture Modeling

- Much slower computations than normal mixtures, especially for large sample sizes
- Needs larger samples; small class sizes can create problems (but successful analyses can be done at $n = 100-200$)
- Needs more random starts than normal mixtures to replicate the best loglikelihood
- Lower entropy
- Needs continuous variables
- Needs continuous variables with many distinct values: Likert scales treated as continuous variables may not carry enough information
- Models requiring numerical integration not yet implemented (required with factors behind categorical and count variables, although maybe not enough information)

Non-normal mixtures

- Can fit the data considerably better than normal mixtures
- Can use a more parsimonious model
- Can reduce the risk of extracting latent classes that are merely due to non-normality of the outcomes
- Can check the stability/reproducibility of a normal mixture solution
- Can describe the percentiles of skewed distributions

Normal mixtures

- Can carve out smaller subgroups in the sample that non-normal mixture modeling might miss, such as tail subgroups with different behaviors
- Can handle smaller class sizes
- Can be computed relatively quickly
- Can be used as a starting point for non-normal mixture modeling
- Can be robust to mild non-normality (LSAY math example, CACE for JOBS data) - skew/kurtosis less than plus/minus 0.5?

6.15 Mplus Specifications

DISTRIBUTION=SKEWT/SKEWNORMAL/TDIST in the ANALYSIS command makes it possible to access non-normality parameters in the MODEL command

- Skew parameters are given as $\{y\}$, $\{f\}$, where the default is $\{f\}$ and class-varying. Having both $\{y\}$ and $\{f\}$ is not identified
- Degrees of freedom parameters are given as $\{df\}$ where the default is class-varying
- $df < 1$: mean not defined, $df < 2$: variance not defined, $df < 3$: skewness not defined. Density can still be obtained
- Class-varying $\{f\}$ makes it natural to specify class-varying f variance
- Normal part of the distribution can get zero variances (fixed automatically), with only the non-normal part remaining

```

VARIABLE:      NAMES = id gender age_1996 age_1997 race1 bmi12_2
                bmi13_2 bmi14_2 bmi15_2 bmi16_2 bmi17_2 bmi18_2 bmi19_2
                bmi20_2 bmi21_2 bmi22_2 bmi23_2 black hisp mixed c1 c2 c3
                c1_wom c2_wom c3_wom momedu par_bmi bio1_bmi bio2_bmi
                bmi_par currsmk97 bingedrnk97 mjuse97 cent_msa
                liv2prnts adopted income hhsz97;
USEVARIABLES = bmi12_2 bmi13_2 bmi14_2
                bmi15_2 bmi16_2 bmi17_2 bmi18_2
                bmi19_2 bmi20_2 bmi21_2 bmi22_2 bmi23_2;
USEOBSERVATIONS = gender EQ -1;
MISSING = ALL (9999);
CLASSES = c(2);
ANALYSIS:      TYPE = MIXTURE;
                STARTS = 400 80;
                PROCESSORS = 8;
DISTRIBUTION = SKEWT;
                ESTIMATOR = MLR;

```

Mplus Input Example, Continued

```
MODEL:          %OVERALL%
               i s q |bmi12_2@-.5 bmi13_2@-.4 bmi14_2@-.3
               bmi15_2@-.2 bmi16_2@-.1 bmi17_2@0 bmi18_2@.1
               bmi19_2@.2 bmi20_2@.3 bmi21_2@.4 bmi22_2@.5 bmi23_2@.6;
               %c#1%
               i-q;
               i-q WITH i-q;
               bmi12_2-bmi23_2(1);
               %c#2%
               i-q;
               i-q WITH i-q;
               bmi12_2-bmi23_2(2);
OUTPUT:        TECH1 TECH4 TECH8 RESIDUAL;
PLOT:         TYPE = PLOT3;
              SERIES = bmi12_2-bmi23_2(s);
```

ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.

THE FOLLOWING PARAMETERS WERE FIXED:

Parameter 5, %C#1%: I
Parameter 6, %C#1%: S WITH I
Parameter 8, %C#1%: Q WITH I

THIS MAY ALSO BE DUE TO RESIDUAL VARIANCES CONVERGING TO 0.
THESE RESIDUAL VARIANCES AND CORRESPONDING COVARIANCES ARE FIXED TO 0.

Mplus Output, Continued

Variances				
I	0.000	0.000	999.000	999.000
S	11.445	3.046	3.758	0.000
Q	44.779	20.085	2.230	0.026
S WITH				
I	0.000	0.000	999.000	999.000
Q WITH				
I	0.000	0.000	999.000	999.000
S	18.823	4.986	3.775	0.000

Skew and Df Parameters

Latent Class 1

I	6.236	0.343	18.175	0.000
S	3.361	0.542	6.204	0.000
Q	-2.746	1.399	-1.963	0.050
DF	3.516	0.403	8.732	0.000

Latent Class 2

I	4.020	0.279	14.408	0.000
S	-0.875	0.381	-2.296	0.022
Q	3.399	1.281	2.653	0.008
DF	3.855	0.562	6.859	0.000

Technical 4 Output: Estimates
derived from the model for Class 1

Estimated means for the
latent variables

	I	S	Q
1	28.138	10.516	-2.567

Estimated covariance matrix for
the latent variables

	I	S	Q
I	48.167		
S	25.959	40.531	
Q	-21.212	32.220	113.186

Estimated correlation matrix
for the latent variables

	I	S	Q
I	1.000		
S	0.588	1.000	
Q	-0.287	0.476	1.000

Estimated skew for the
latent variables

	I	S	Q
1	6.653	3.437	-1.588

7. SEM Allowing Non-Normal Distributions

Non-Normal SEM with t, skew-normal, and skew-t distributions:

- Allowing a more general model, including non-linear conditional expectation functions and heteroscedasticity
- Chi-square test of model fit using information on skew and df
- Missing data handling avoiding the normality assumption of FIML
- Mediation modeling allowing general direct and indirect effects
- Percentile estimation of the skewed factor distributions

Mplus offers six types of SEM:

- Regular SEM: Distribution = Normal
- ESEM: Exploratory factor analysis measurement model
- BSEM: Bayesian analysis with small-variance priors
- skew-SEM: Non-normal SEM; Distribution = tdist/skew/skewt

- ML estimate robustness to non-normality in SEM
- SEs and chi-square can be adjusted for non-normality (sandwich estimator, "Satorra-Bentler")
- MLE robustness doesn't hold if residuals and factors are not independent (Satorra, 2002) as with residual skew $\neq 0$
- Asparouhov-Muthén (2014):

There is a preconceived notion that standard structural models are sufficient as long as the standard errors of the parameter estimates are adjusted for failure of the normality assumption, but this is not really correct. Even with robust estimation the data is reduced to means and covariances. Only the standard errors of the parameter estimates extract additional information from the data. The parameter estimates themselves remain the same, i.e., the structural model is still concerned with fitting only the means and the covariances and ignoring everything else.

$$Y = \nu + \Lambda\eta + \varepsilon$$

$$\eta = \alpha + B\eta + \Gamma X + \xi$$

where

$$(\varepsilon, \xi) \sim rMST(0, \Sigma_0, \delta, DF)$$

and

$$\Sigma_0 = \begin{pmatrix} \Theta & 0 \\ 0 & \Psi \end{pmatrix}.$$

The vector of parameters δ is of size $P + M$ and can be decomposed as $\delta = (\delta_Y, \delta_\eta)$. This gives the conditional distributions

$$\eta|X \sim rMST((I - B)^{-1}(\alpha + \Gamma X), (I - B)^{-1}\Psi((I - B)^{-1})^T, (I - B)^{-1}\delta_\eta, DF),$$

$$Y|X \sim rMST(\mu, \Sigma, \delta_2, DF), \quad (32)$$

$$\mu = \nu + \Lambda(I - B)^{-1}(\alpha + \Gamma X), \quad (33)$$

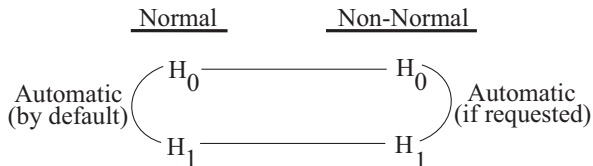
$$\Sigma = \Theta + \Lambda(I - B)^{-1}\Psi((I - B)^{-1})^T\Lambda^T, \quad (34)$$

$$\delta_2 = \delta_Y + \Lambda(I - B)^{-1}\delta_\eta. \quad (35)$$

7.1 SEM Chi-Square Testing with Non-Normal Distributions

Adding skew and df parameters to the means, variances, and covariances of the unrestricted H1 model

SEM Likelihood-Ratio Chi-Square Testing



- Non-normal H0 vs H1 test obtained by the H1MODEL option of the OUTPUT command
- Not provided by default - computationally demanding
- H1MODEL has two settings: COVARIANCE (default) and SEQUENTIAL
- H1STARTS: Needed because convergence may be difficult for the H1 model

7.2 Missing Data with Non-Normal Distributions

- Modeling with missing data via the FIML estimator is not robust to the normality assumptions
- The t-distribution EM-algorithm by Liu and Rubin (1995)

A simulated example of the effect of normality violation:

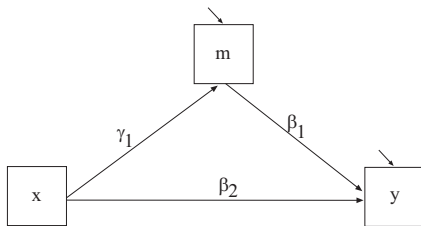
- Skew-normal distribution with 5 variables and $n = 100000$
- Missing data for the first variable using the following MAR missing data mechanism

$$P(Y_1 \text{ is missing}) = \frac{1}{1 + \text{Exp}(-1 + Y_2 + Y_3 + Y_4 + Y_5)}. \quad (36)$$

- Population mean for Y_1 is $3\sqrt{2/\pi} \approx 2.4$
- The standard FIML estimator assuming normality estimates the mean of Y_1 as 2.1
- Using the correct distributional assumption and estimating the saturated skew-normal model the mean estimate is 2.4

7.3 Path Analysis of Firefighter Data ($n = 354$)

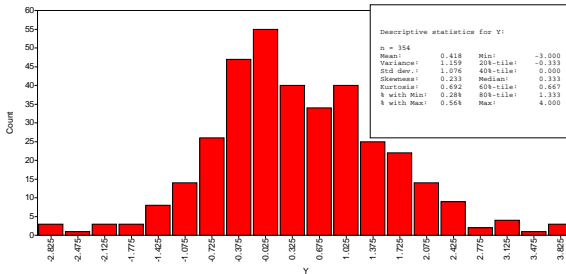
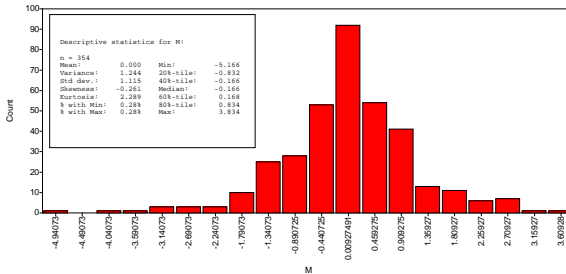
Intervention on eating fruits and vegetables: M= knowledge, Y = eating.



Sources:

- Yuan & MacKinnon (2009). Bayesian mediation analysis. Psychological Methods
- Elliot et al. (2007). The PHLAME (Promoting Healthy Lifestyles: Alternative Models Effects) firefighter study: outcomes of two models of behavior change. Journal of Occupational and Environmental Medicine

Firefighter M and Y Distributions



Summary Comparison of Models for the Firefighter Example

Distribution	LL	Number of parameters	BIC
Normal	-1058	7	2157
t-dist	-1045	8 (adding df)	2137
Skew-normal	-1055	9 (adding 2 skew)	2162
Skew-t	-1043	10 (adding df and 2 skew)	2144

- The df parameter of the t-distribution is needed to capture the kurtosis, but skew parameters are not needed
- The t-distribution allows for heteroscedasticity in the Y residual as a function of M; the conditional expectation functions are linear; usual indirect effect valid

The skew-normal run needed STARTS= 32 8

Comparing Normal and T-Distribution Estimates

Normal distribution - regular SEM				
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
M O N				
X	0.397	0.119	3.346	0.001
Y O N				
M	0.142	0.051	2.755	0.006
X	0.108	0.116	0.926	0.354
Intercepts				
Y	0.418	0.056	7.417	0.000
M	0.000	0.058	0.000	1.000
Residual Variances				
Y	1.125	0.085	13.304	0.000
M	1.203	0.090	13.304	0.000
New/Additional Parameters				
INDIRECT	0.056	0.026	2.127	0.033
T-distribution				
M O N				
X	0.371	0.110	3.384	0.001
Y O N				
M	0.119	0.059	2.003	0.045
X	0.134	0.115	1.161	0.246
Intercepts				
Y	0.384	0.055	6.940	0.000
M	0.005	0.053	0.093	0.926
Residual Variances				
Y	0.872	0.088	9.963	0.000
M	0.829	0.092	9.006	0.000
Skew and Df Parameters				
DF	7.248	1.851	3.915	0.000
New/Additional Parameters				
INDIRECT	0.044	0.026	1.718	0.086

What Makes the Normal and t-Distribution Results Different?

Liu & Rubin (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica*, 5, 19-39.

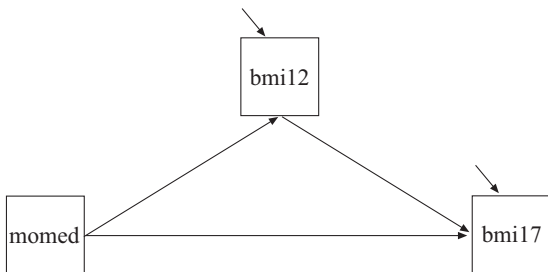
Consider the mean of M in the two treatment groups (effect of X on M). Under the t-distribution they are estimated by weighting by (see Liu-Rubin, eqn (23))

$$(df + 1)/(df + MD), \quad (37)$$

where MD is the Mahalanobis distance, that is, the t-distribution protects against outliers.

	Normal Mean	t-distribution Mean
X=0	-0.163 (.069)	-0.124 (.062)
X=1	0.234 (.102)	0.194 (.093)
Difference	0.397 (.123)	0.318 (.114)

7.4 Path Analysis Mediation Model for BMI



bmi12 skewness/kurtosis = 1.34/2.77

bmi17 skewness/kurtosis = 1.86/5.29

BMI Mediation Model: Normal Distribution Results using MLR

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
BMI17 ON				
BMI12	0.939	0.054	17.307	0.000
MOMED	0.021	0.096	0.221	0.825
BMI12 ON				
MOMED	-0.430	0.098	-4.394	0.000
Intercepts				
BMI12	20.891	0.287	72.670	0.000
BMI17	4.811	1.126	4.272	0.000
Residual Variances				
BMI12	15.484	1.392	11.127	0.000
BMI17	12.357	1.417	8.720	0.000
New/Additional Parameters				
INDIRECT	-0.403	0.094	-4.269	0.000

Summary Comparison of Models for the BMI Example

Distribution	LogL	Number of parameters	BIC
Normal	-11207	7	22471
t-dist	-10789	8 (adding df)	21642
Skew-normal	-10611	9 (adding 2 skew)	21295
Skew-t	-10423	10 (adding df and 2 skew)	20928

The key formula for counterfactual effects is

$$E[Y(x, M(x^*))] = \int_{-\infty}^{+\infty} E[Y|X = x, M = m] \times f(m|X = x^*) \partial m. \quad (38)$$

Asparouhov-Muthén (2014): The above integral is the conditional Y mean of the skew-t distribution $rMST(\mu, \sigma, \delta, \nu)$ as in (32) on slide 176, where

$$\mu = \alpha_1 + \beta_1(\alpha_2 + \beta_3 x^*) + \beta_2 x, \quad (39)$$

$$\sigma = \sigma_1 + \beta_1^2 \sigma_2, \quad (40)$$

$$\delta = \delta_1 + \beta_1 \delta_2. \quad (41)$$

Using the formula for the mean of a skew t-distribution we obtain

$$E[Y(x, M(x^*))] = \alpha_1 + \beta_1(\alpha_2 + \beta_3 x^*) + \beta_2 x + (\delta_1 + \beta_1 \delta_2) \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \sqrt{\frac{\nu}{\pi}}. \quad (42)$$

Counterfactual Effects with the Skew T-Distribution, Continued

The causal direct and indirect effects are computed from

$$E[Y(x_1, M(x_1^*))] - E[Y(x_2, M(x_2^*))] = \beta_1 \beta_3 (x_1^* - x_2^*) + \beta_2 (x_1 - x_2). \quad (43)$$

In line with VanderWeele and Vansteelandt (2009), special cases of this formula give the direct effect as

$$E[Y(x, M(x^*))] - E[Y(x^*, M(x^*))] = \beta_2 (x - x^*) \quad (44)$$

and the indirect effect as

$$E[Y(x, M(x)) - E[Y(x, M(x^*))]] = \beta_1 \beta_3 (x - x^*). \quad (45)$$

The above formulas are identical to the normal distribution case and thus the direct and indirect effects are not affected by the skewness of the residuals.

BMI Mediation Model: Input for Skew t-Distribution

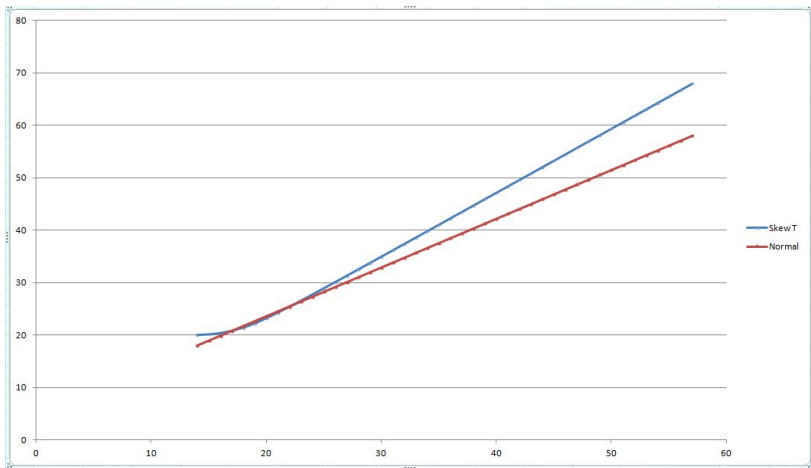
```
DATA: FILE = 2.dat;
VARIABLE: NAMES = bmi12 bmi17 momed;
MISSING = ALL(999);
ANALYSIS: DISTRIBUTION = SKEWT;
PROCESSORS = 8;
MODEL: bmi17 ON bmi12 (b)
momed;
bmi12 ON momed (a);
MODEL CONSTRAINT: NEW(indirect);
indirect = a*b;
OUTPUT: TECH1 TECH4 TECH8 RESIDUAL STANDARDIZED;
PLOT: TYPE = PLOT3;
```

BMI Mediation Model: Results for Skew t-Distribution

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
BMI17 ON				
BMI12	0.226	0.076	2.971	0.003
MOMED	-0.046	0.042	-1.088	0.276
BMI12.2 ON				
MOMED	-0.239	0.077	-3.087	0.002
Intercepts				
BMI12	17.424	0.309	56.315	0.000
BMI17	14.832	1.356	10.939	0.000
Residual Variances				
BMI12	4.067	0.568	7.165	0.000
BMI17	1.377	0.200	6.880	0.000
Skew and Df Parameters				
BMI12	2.876	0.259	11.088	0.000
BMI17	4.347	0.255	17.022	0.000
DF	3.851	0.283	13.587	0.000
New/Additional Parameters				
INDIRECT	-0.054	0.027	-1.974	0.048

Regression of BMI17 on BMI12: Skew-T vs Normal



A Closer Look at the Skew-T Model: Allowing for Varying Skews and Variances

The skews and residual variances are not constant over the X (Momed) values for the BMI example. Asparouhov-Muthén propose

$$BMI12 = \alpha_1 + \varepsilon_1, \quad (46)$$

$$BMI17 = \alpha_2 + \beta_1 BMI12 + \varepsilon_2, \quad (47)$$

where with the notation $\varepsilon \sim rMST(\mu, \sigma, \delta, \nu)$ linearity is specified

$$\varepsilon_1 \sim rMST(0, a_1 + b_1 X, a_2 + b_2 X, \nu), \quad (48)$$

$$\varepsilon_2 \sim rMST(0, a_3 + b_3 X, a_4 + b_4 X, \nu), \quad (49)$$

where X represents the mother's education predictor.

- Initial skew-t model BIC = 30849
- Modified skew-t model BIC = 30805 (a and b coefficients all negative, significant: Increasing X gives a more normal and homogenous BMI distribution)

Using the counterfactual effect formula Asparouhov-Muthén (2014) show that

$$E[Y(x, M(x^*))] = \alpha_2 + \beta_1 \alpha_1 + \left(\beta_1 (a_2 + b_2 x^*) + a_4 + b_4 x \right) \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \sqrt{\frac{\nu}{\pi}}, \quad (50)$$

resulting in the direct effect

$$E[Y(x, M(x^*))] - E[Y(x^*, M(x^*))] = b_4 (x - x^*) \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \sqrt{\frac{\nu}{\pi}} \quad (51)$$

and the indirect effect

$$E[Y(x, M(x))] - E[Y(x, M(x^*))] = \beta_1 b_2 (x - x^*) \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \sqrt{\frac{\nu}{\pi}}. \quad (52)$$

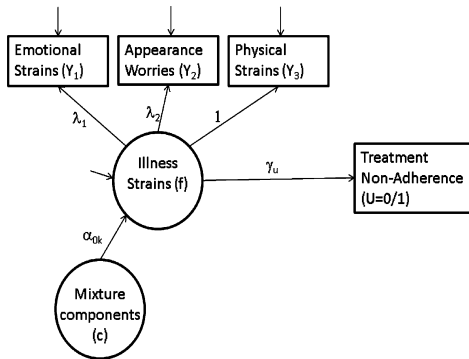
- The modified skew-t model results indicate that the direct effect is 85% of the total effect while the indirect effect is only 15% of the total effect
- The normal model found that the direct effect was ignorable relative to the indirect effect
- These drastically different results illustrate the modeling opportunities when we look beyond the mean and variance modeling used with standard SEM

7.5 Non-Normal CFA

- Wall, Guo, & Amemiya (2012). Mixture factor analysis for approximating a nonnormally distributed continuous latent factor with continuous and dichotomous observed variables.

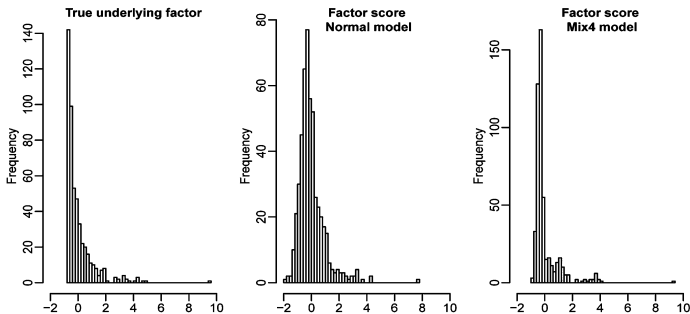
Multivariate Behavioral Research

- Normal-ML estimates robust to non-normality for λ s, but not for γ_u



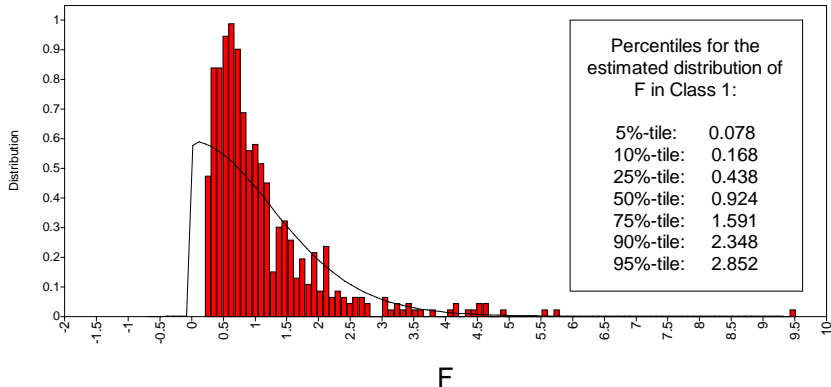
(c)

Figure 6 of Wall et al. (2012):



- Factor distribution can be more parsimoniously specified as skew-t than the authors' mixture of normals
- Mplus gives percentiles for the estimated distribution

Factor Distribution Estimated by 1-Class Skew-T

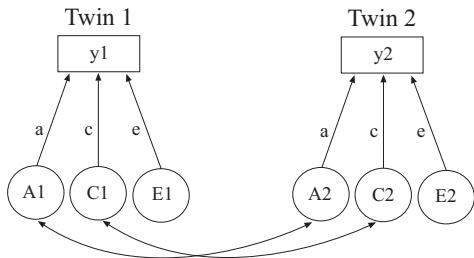


Skewness in Factor Indicators not Explained by the Factor Skewness

Section 4.6 of Asparouhov-Muthén (2014) considers an example with a skew-t factor influencing a set of indicators, one of which has a skew-t residual, and where the factor is regressed on a covariate.

- The skew-t model gives correct chi-square behavior and unbiased estimates
- The presumed non-normality robust normal theory estimator (MLR) chi-square rejects much too often; factor indicator residual not independent of factor
- The normal-theory estimator does not capture the non-linear relationship for the factor indicator regressed on the factor
- The normal-theory estimator gives a biased regression coefficient for the factor regressed on the covariate

ACE Twin Modeling using Skew-T



- How will heritability estimates be affected?
- For which variables should the skew parameters be applied?
- How should growth mixtures be incorporated: Different heritability for different trajectory classes; is there genetic influence on class membership? Skew-t extensions of Muthén, Asparouhov & Rebollo (2006): Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. Twin Research and Human Genetics.

The End

Papers, videos, and handouts can be found at www.statmodel.com