Using Mplus To Investigate Direct Effects in Latent Class Analysis

Bengt Muthén
Professor Emeritus, UCLA
Mplus: `https://www.statmodel.com`
bmuthen@statmodel.com

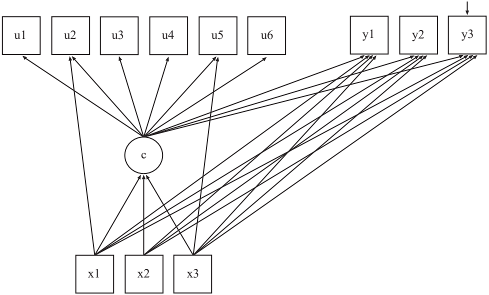
Tihomir Asparouhov
Mplus

Mplus Web Talks: No. 9
July 2025

We thank Thuy Nguyen and Noah Hastings for expert assistance.

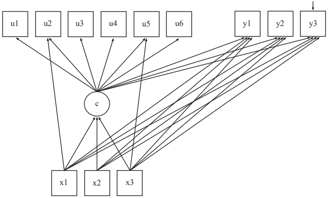
Outline

Direct Effects from Covariates to Latent Class Indicators



- Two direct effects: From x1 to u2 and from x3 to u5
- A direct effect implies that for a given class, the probability of u is not the same for different values of x
  - The measurement model parameters are not the same for different individuals - referred to as measurement non-invariance, differential item functioning, item bias
  - For example, with a binary x describing two groups, the measurement instrument does not work the same for the groups

Direct Effects: Analysis Impact



- Consequences of ignoring direct effects:
  - Violation of conditional independence given C
    - Class enumeration for measurement model - impact likely small
  - Distorted model estimates:
    - Class probabilities - typically small impact (with or without X)
    - C ON X - large impact (direct effects are forced to go through only C (Asparouhov & Muthén, 2014; Web Note 15)
    - Y ON X - some impact (when C ON X changes, the indirect effects of X on Y via C change and therefore Y ON X)
- Multistep analysis
  - Let the measurement model include the X part and its direct effects on the latent class indicators, then add distal outcomes

Outline

- Background
- **Antisocial behavior example**
- Searching for direct effects
  - Different approaches, Penalized SEM (PSEM)
  - ASB example (reduced version)
    - Latent classes and C ON X results
- Evaluating direct effects
  - Direct effects results, calculator
- Class-varying direct effects
  - PSEM DIFF priors, ALIGN option
- Multistep analysis with direct effects and distal outcomes
- Recap and further research

Direct Effects: Antisocial Behavior Example

- ASB data:
  - 17 antisocial behavior items collected in the 1980 National Longitudinal Survey of Youth for respondents between the ages of 16 and 23 together with a set of background variables
  - The ASB items assessed the frequency of various behaviors during the past year, here dichotomized as 0 vs > 0 times
  - A sample of 7,326 respondents has complete data on the antisocial behavior items and the background variables
- ASB analyses:
  - SEM (MIMIC) 4-factor analysis (Muthén, 2025)
  - Latent class analysis from Mplus Short Course Topic 5, slides 91-118
    - 4-class and 5-class LCA of the 17 latent class indicators
    - 5 classes: High, property offense, drug, person offense, normative (low, except for pot)
    - 4 classes used in this talk
  - ASB is a general population survey so that considerable heterogeneity among individuals can be expected - direct effects
  - 17U, 11X version and 7U, 7X reduced version with 2 distals

Input for ASB Analysis with 17 U's and 11 X's C ON X but No Direct Effects

```
TITLE:
DATA:       FILE = asbfree.dat;
            FORMAT = 34X 54F2.0;
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
            sex black hisp single divorce dropout
            college onset age94 dep abuse;
            CATEGORICAL = property-gambling;
            CLASSES = c(4);
ANALYSIS:   TYPE = MIXTURE;
            ESTIMATOR = ML;
            STARTS = 400 100;
            PROCESSORS = 12;
MODEL:      %OVERALL%
            c ON sex-abuse;
OUTPUT:     TECH10 SVALUES;
```

- 187 possible direct effects - which ones are important to include?

Outline

- Background
- Antisocial behavior example
- **Searching for direct effects**
  - **Different approaches, Penalized SEM (PSEM)**
  - ASB example (reduced version)
    - Latent classes and C ON X results
- Evaluating direct effects
  - Direct effects results, calculator
- Class-varying direct effects
  - PSEM DIFF priors, ALIGN option
- Multistep analysis with direct effects and distal outcomes
- Recap and further research

Approaches for Searching for Direct Effects

- Analysis with latent class variable regressed on all covariates and in addition:
  - Each latent class indicator regressed on all covariates or All latent class indicators regressed on one covariate
    - Pro: Each analysis easily converges
    - Con: Several analyses - one for each latent class indicator/covariate
  - All latent class indicators regressed on all covariates
    - Pro: Single analysis
    - Con: May not converge or be empirically identified
      - Relies on higher-order moments (cf. the non-identified case of MIMIC with direct effects for continuous factor indicators)
  - All latent class indicators regressed on all covariates using PSEM regularization, Asparouhov & Muthén (2024).
    - PSEM for mixtures: Asparouhov & Muthén (2025)
    - Pro: Leads to parsimonious models, i.e., fewer significant direct effects, converges more easily for a small-enough prior variance
    - Con: Prior variance choice calls for more than one analysis

PSEM Applied to Finding Direct Effects in LCA

- PSEM: Penalized structural equation modeling
  - ML estimation using priors
- PSEM uses priors for two purposes:
  - Estimating models that are non-identified without the priors
    - Similar to BSEM for Bayes estimation
  - Simplifying models that are identified but can be fitted practically as well with fewer parameters
    - Reguralized analysis (RegSEM). Common prior: LASSO
    - Example: Direct effects in LCA
- Including all direct effects, PSEM uses the LASSO or ALF (Alignment Loss Function) mean and variance priors for the directs effects,

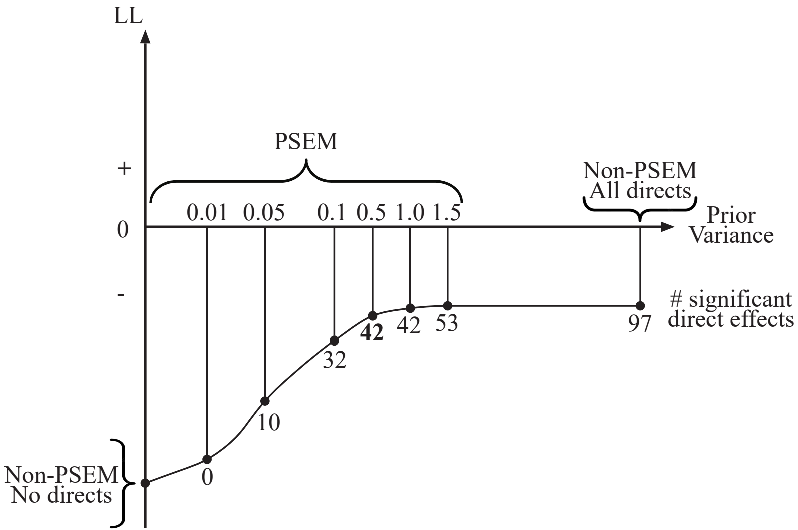
  maximizing: fit function = log likelihood + penalty

  where the penalty is larger for smaller variance. Ex: ALF(0, 0.01)
  - The penalty (which is negative) penalizes models with many direct effects - favors a parsimonious model
    - Variance = 0: Same as non-PSEM analysis with no direct effects
    - Variance = ∞: Same as non-PSEM analysis with all direct effects
  - Goal: use a variance that makes the logL practically as good as with all direct effects included - but with fewer direct effects

Input for PSEM with ALF(0, 0.5): 17 U's, 11 X's

```
TITLE:      17 U's, 11 X's,
            PSEM ALF(0.5)
            Model 6 on slide 12

DATA:       FILE = asbfree.dat;

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;

            USEVAR = property-gambling
            sex black hisp single divorce dropout
            college onset age94 dep abuse;

            CATEGORICAL = property-gambling;
            CLASSES = c(4);

ANALYSIS:   TYPE = MIXTURE;
            ESTIMATOR - ML;
            STARTS = 800 200;
            PROCESSORS = 12;

MODEL:      %OVERALL%
            c ON sex-abuse;
            ! Direct effects:
            property-gambling ON
            sex-abuse (d1-d187);

MODEL PRIOR: d1-d187~ALF(0,0.5);

OUTPUT:     TECH1 TECH3 TECH10
            SVALUES;
```

ASB Log Likelihoods: 17 Us, 11 Xs (187 Possible Effects)



ASB Log Likelihood and BIC Values for 17 Us, 11 Xs

Model	# par's	logL	BIC	#sig. dir.	logL drop
1. No directs	104	-40,088.255	81,102	0	
2. All directs	291	-39,278.105	81,146	97	
3. 97 sig. directs	**201**	**-39,388.584**	**80,466**	**97**	
4. PSEM (0.05)	NA	-39,564.585	NA	10	
5. PSEM (0.1)	NA	-39,384.232	NA	32	
6. **PSEM (0.5)**	NA	-39,295.792	NA	42	
7. **PSEM (1.0)**	NA	-39,282.084	NA	42	
8. PSEM (1.5)	NA	-39,279.840	NA	53	
9. 4: 10 directs	114	-39,855.419	80,725	10	1.2 %
10. 5: 32 directs	136	-39,531.593	80,273	32	0.4 %
11. 6: 42 directs	**146**	**-39,466.120**	**80,232**	**41**	0.2 %
12. 7: 42* directs	**146**	**-39,485.213**	**80,270**	**41**	0.2 %
13. 8: 53 directs	157	-39,483.475	80,364	52	0.2 %

- logL drop is computed as the percentage $100 * (logL - logL_{Model\,3})/logL_{Model\,3}$
- * The 42 direct effects are not all the same in models 11 and 12

Discussion of ASB Models for 17 Us, 11 Xs: Why is 42 for PSEM(0.5) bolded in the Graph of Slide 12?

- PSEM(0.5) model 6 has a lower (worse) logL than PSEM(1.0) model 7
- Both models show 42 significant direct effects - but not the same ones
- The non-PSEM models 11 and 12 are based on PSEM models 6 and 7
  - Freeing the 42 effects results in a higher logL for model 11 than for model 12 despite model 6 having a lower logL than model 7
- Models 11 and 12 have the same
  - Number of parameters
  - Number of significant direct effects
  - LogL 0.2 % drop
- Model 11 is chosen
  - Better logL and BIC than model 12
  - Best BIC of all the models

Outline

- Background
- Antisocial behavior example
- Searching for direct effects
  - Different approaches, Penalized SEM (PSEM)
  - **ASB example (reduced version)**
    - Latent classes and C ON X results
- Evaluating direct effects
  - Direct effects results, calculator
- Class-varying direct effects
  - PSEM DIFF priors, ALIGN option
- Multistep analysis with direct effects and distal outcomes
- Recap and further research

ASB Inputs for the Reduced Set of 7 U's, 7 X's DATA and VARIABLE Commands for All Runs

```
TITLE:

DATA:        FILE = asbfree.dat;

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 fight shoplift threat
             pot drug goods
             sex black hisp single divorce dropout age94;

             CATEGORICAL = property-goods;
             CLASSES = c(4);
```

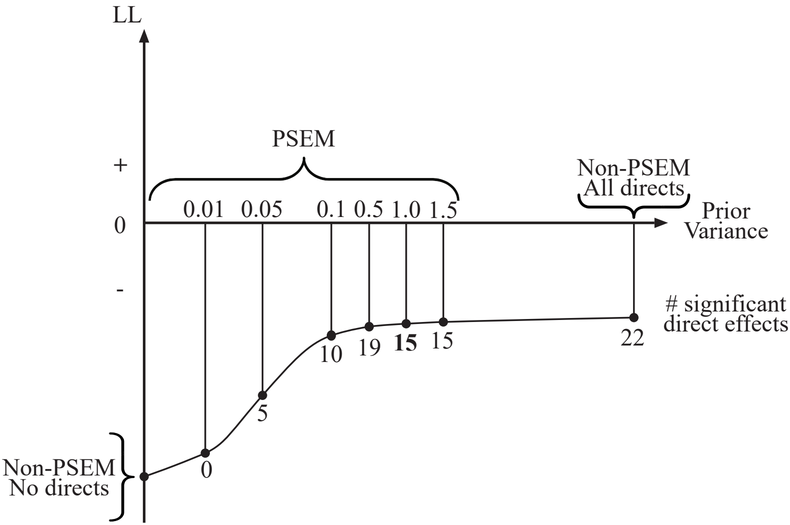
All U's on all X's

```
ANALYSIS:     TYPE = MIXTURE;
              ESTIMATOR = ML;
              STARTS = 400 100;
              PROCESSORS = 8;

MODEL:        %OVERALL%
              c ON sex-age94;
              property-goods ON sex-age94;

OUTPUT:       TECH10 SVALUES;
```

All U's on all X's, PSEM (1.0)

```
ANALYSIS:     TYPE = MIXTURE;
              ESTIMATOR = ML;
              STARTS = 400 100;
              PROCESSORS = 8;

MODEL:        %OVERALL%
              c ON sex-age94;
              property-goods ON
              sex-age94 (d1-d49);

OUTPUT:       TECH10 SVALUES;

MODEL PRIOR:  d1-d49 ~ ALF(0,1.0);
              ! Computing time: 1 minute
              ! (17U-11X run takes 11 minutes)
```

ASB Log Likelihoods: 7 Us, 7 Xs (49 Possible Effects)



ASB Log Likelihood and BIC Values for 7 Us, 7 Xs

Model	# par's	logL	BIC	#sig. dir.	logL drop
1. No directs	52	-23,930.193	48,323	0	
2. All directs	101	-23,653.011	48,205	22	
3. 22 sig. directs	**74**	**-23,670.477**	**47,999**	**22**	
4. PSEM (0.1)	NA	-23,680.907	NA	10	
5. PSEM (0.5)	NA	-23,656.804	NA	19	
6. **PSEM (1.0)**	NA	-23,654.305	NA	15	
7. PSEM (1.5)	NA	-23,653.627	NA	15*	
8. 4: 10 directs	62	-23,774.174	48,100	10	0.4 %
9. 5: 19 directs	71	-23,683.270	47,998	19	0.1 %
10. 6: 15 directs	**67**	**-23,715.072**	**48,026**	**14**	0.2 %

- logL drop is computed as the percentage $100 * (logL - logL_{Model\,3})/logL_{Model\,3}$
- * The 15 direct effects are the same in models 6 and 7

Discussion of ASB Models for 7 Us, 7 Xs: Why is 15 for PSEM(1.0) bolded in the Graph of Slide 18?

- PSEM(1.0) model 6 and PSEM(1.5) model 7 have the same number of significant direct effects
  - It doesn't matter that model 7 has a higher logL because both models lead to the non-PSEM model 10
- Model 10 is chosen because it has a small 0.2 % drop in logL, is relatively parsimonious, and is close to the best BIC
  - Model 9 is a strong contender with a smaller 0.1 % drop in logL, the best BIC, but 19 instead of 15 direct effects to consider
  - Model 3 is also a contender with almost the same BIC as model 9, but has 22 instead of 15 direct effects
    - Are the extra direct effects of model 3 and model 9 important? [1]
- Class-specific direct effects can be explored based on model 10 using PSEM DIFF priors

[1] For model 9, Fight on Dropout is the only noteworthy direct effect beyond those of model 10 (probability difference 0.14 for non-hispanic males at average Age94)

Technical Notes

- # parameters = NA because the RegSEM version of PSEM doesn't have a good way of counting the parameters:
  - Parameters can have tiny prior variances with parameters estimated close to 0, and contributing nothing to fit
    - These are not real parameters and shouldn't be counted, but the precise removal of these parameters is subjective since it will need a definition of how small is really 0
- Alternative log likelihood drop % definitions:
  - PSEM models: (L - L(M2)) / (L(M1) - L(M2))
    - The logic is that the H1 model here is M2 and baseline is M1
  - Models based on PSEM: (L - L(M3)) / (L(M1) - L(M3))
    - The logic is that the H1 model here is M3 and baseline is M1
  - % drop relative to the total possible drop: Cut-off $\leq 5\%$?
  - For the full set of variables, model 11 does best - same model choice as on slides 13-14
  - For the reduced set of variables, model 9 does best with 19 direct effects (5% drop), whereas slides 18-20 chose model 10 with 15 direct effects (17% drop) - are the extra direct effects important?

Outline

- Background
- Antisocial behavior example
- Searching for direct effects
  - Different approaches, Penalized SEM (PSEM)
  - ASB example (reduced version)
    - **Latent classes and C ON X results**
- Evaluating direct effects
  - Direct effects results, calculator
- Class-varying direct effects
  - PSEM DIFF priors, ALIGN option
- Multistep analysis with direct effects and distal outcomes
- Recap and further research

Interpreting the 4 Latent Classes for Model 10 with 15 Direct Effects

- Classes (probability):
  - Fight/Threat (0.321), High (0.130), Drugs (0.172), Low (0.377)
- The output section RESULTS IN PROBABILITY SCALE provides convenient interpretation, especially with binary latent class indicators
  - Probability of being in the high category of the indicator:

Variable	Class 1	Class 2	Class 3	Class 4
PROPERTY	0.18	0.72	0.14	0.03
FIGHT	**0.49**	0.73	0.12	0.02
SHOPLIFT	0.22	0.79	0.39	0.07
THREAT	**0.58**	0.80	0.30	0.05
POT	0.32	0.90	**0.96**	0.18
DRUG	0.03	0.53	**0.55**	0.01
GOODS	0.08	0.53	0.09	0.01

Class Probabilities for 7 U's on 7 X's Comparing Three Models

- 4 classes: Fight/Threat, High, Drugs, Low
  - Same interpretation for these 3 different models
- Class probabilities with 15 direct effects:
  0.321, 0.130, 0.172, 0.377
- Class probabilities with no direct effects:
  0.298, 0.126, 0.189, 0.386
- Class probabilities with latent class indicators only:
  0.219, 0.106, 0.209, 0.466

Comparing C ON X Results for 7 U's on 7 X's: 15 Directs (Model 10) vs No Directs for Class 1

15 direct effects, Model 10

C#1 ON X

	Estimate	S.E.	Est./S.E.
SEX	1.286	0.102	12.627
BLACK	1.261	0.128	9.834
HISP	**-0.256**	0.128	**-1.993**
SINGLE	0.011	0.107	0.105
DIVORCE	0.272	0.120	**2.278**
DROPOUT	0.282	0.132	2.139
AGE94	-0.204	0.024	-8.634

No direct effects

C#1 ON X

	Estimate	S.E.	Est./S.E.
SEX	1.590	0.101	15.809
BLACK	1.088	0.114	9.572
HISP	**0.210**	0.121	**1.734**
SINGLE	0.068	0.109	0.620
DIVORCE	0.141	0.122	**1.160**
DROPOUT	0.427	0.131	3.252
AGE94	-0.316	0.025	-12.673

- Number of significant effects of covariates on all the latent classes:
  - 15 direct effects model: 17
  - No direct effects model: 13
- Similar discrepancies found in the simulations of Asparouhov & Muthén (2014; Web Note 15)

Outline

- Background
- Antisocial behavior example
- Searching for direct effects
    - Different approaches, Penalized SEM (PSEM)
    - ASB example (reduced version)
        - Latent classes and C ON X results
- **Evaluating direct effects**
    - Direct effects results, calculator
- Class-varying direct effects
    - PSEM DIFF priors, ALIGN option
- Multistep analysis with direct effects and distal outcomes
- Recap and further research

15 Direct Effects for 7 U's on 7 X's (Model 10)

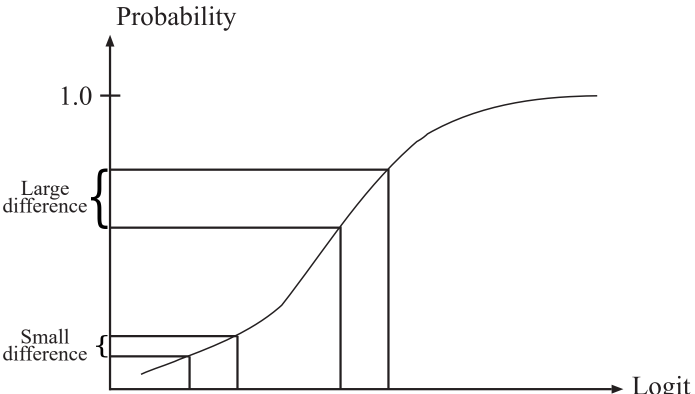
- Property: Sex, Black, Divorce, Age94
- Fight: Sex, Hisp, Age94
- Shoplift: Black, Hisp
- Threat: Black
- Pot: Age94
- Drug: Black, Age94
- Goods: Sex, Black

Direct Effect Estimates for Property and Fight Indicators

	Estimate	S.E.	Est./S.E.
PROPERTY ON			
SEX	**0.676**	0.098	6.925
BLACK	-0.744	0.111	-6.673
DIVORCE	-0.458	0.112	-4.078
AGE94	-0.100	0.021	-4.857
FIGHT ON			
SEX	**0.587**	0.090	6.501
HISP	0.731	0.125	5.860
AGE94	-0.110	0.019	-5.658

- The sex (male) effect is larger for Property than for Fight
- But the effect on their probabilities also depends on their thresholds

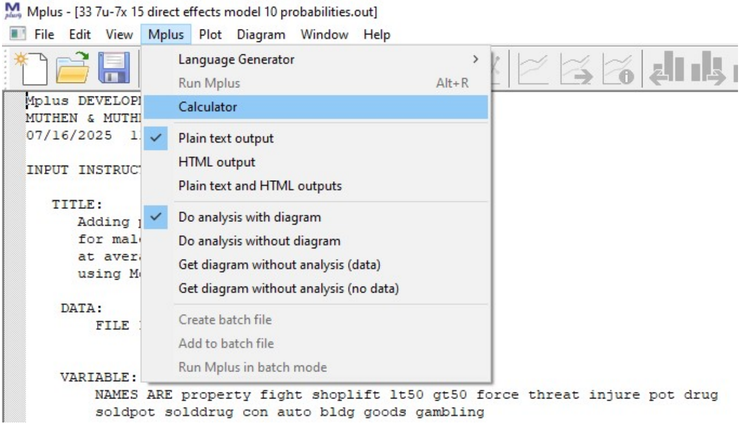
Direct Effect Probabilities: Prob = $1/(1+e^{-Logit})$

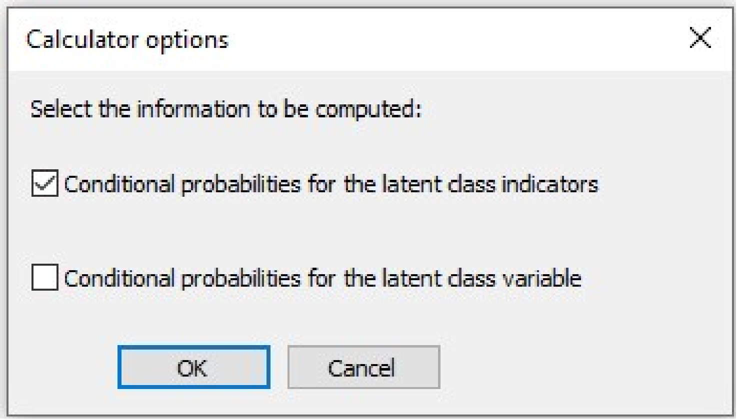


- Logit = - $\tau_c + \beta\ x$, where τ_c is a threshold for class c and β is the direct effect
- Small probability obtained with large threshold, resulting in small logit
- Largest probability change from $x = 0$ to $x = 1$ occurs at the steepest part of the curve with threshold close to zero, that is, probability close to 0.5 (logit = 0 gives probability = 0.5)

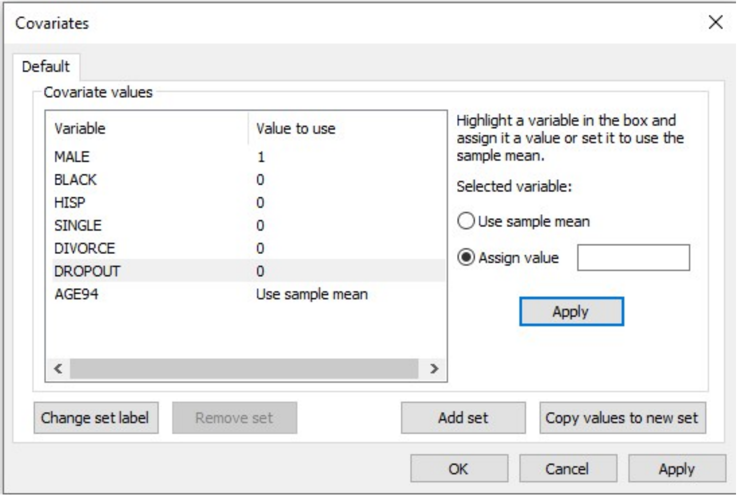
Direct Effect Probabilities: Property and Fight Indicators Comparing Males and Females for the Fight/Threat Class

- Model 10 estimates:
  - Property:
    - Threshold = 1.352, β_{male} = 0.676 (Logit = -0.68 when other x's=0)
  - Fight:
    - Threshold = 0.217, β_{male} = 0.587 (Logit = 0.37 when other x's=0)
- Probabilities for females (sex=0) vs males (sex=1) at zero values for all other covariates, except age94 which is at its sample mean:
  - Property: 0.161 vs 0.274 (difference = 0.113)
  - Fight: 0.368 vs 0.511 (difference = 0.143)
- Compared to the Property indicator, the Fight indicator has a smaller threshold and a smaller direct effect slope and therefore a logit closer to zero with a probability closer to 0.5 where the probability curve is steeper, resulting in a larger probability difference for Fight than Property - the direct effect slope size alone does not tell the whole story


Mplus - [33 7u-7x 15 direct effects model 10 probabilities.out]
File Edit View Mplus Plot Diagram Window Help
Language Generator
Run Mplus Alt+R
Calculator
Plain text output
HTML output
Plain text and HTML outputs
Do analysis with diagram
Do analysis without diagram
Get diagram without analysis (data)
Get diagram without analysis (no data)
Create batch file
Add to batch file
Run Mplus in batch mode
Mplus DEVELOP
MUTHEN & MUTH
07/16/2025 1
INPUT INSTRUC
TITLE:
Adding
for mal
at aver
using M
DATA:
FILE
VARIABLE:
NAMES ARE property fight shoplift lt50 gt50 force threat injure pot drug
soldpot solddrug con auto bldg goods gambling


Calculator options
Select the information to be computed:
Conditional probabilities for the latent class indicators
Conditional probabilities for the latent class variable
OK
Cancel

Calculator Settings: Male=1



Calculator Results: Male=1

Results of the calculator computations

```
Results for the condition(s):
        MALE = 1.000000
        BLACK = 0.000000
        HISP = 0.000000
        SINGLE = 0.000000
        DIVORCE = 0.000000
        DROPOUT = 0.000000
        All other covariates set at sample mean

CONDITIONAL PROBABILITIES FOR THE LATENT CLASS INDICATORS

          PROBABILITY PROFILES
                Latent Class
                   1        2        3        4
PROPERTY
     Category 1  0.726    0.189    0.780    0.952
     Category 2  0.274    0.811    0.220    0.048
FIGHT
     Category 1  0.489    0.270    0.858    0.975
     Category 2  0.511    0.730    0.142    0.025
SHOPLIFT
     Category 1  0.785    0.211    0.622    0.940
     Category 2  0.215    0.789    0.378    0.060
```

Copy | Save to file | Close

Calculator Results: Male=0

Results of the calculator computations

```
Results for the condition(s):
        MALE = 0.000000
        BLACK = 0.000000
        HISP = 0.000000
        SINGLE = 0.000000
        DIVORCE = 0.000000
        DROPOUT = 0.000000
        All other covariates set at sample mean

CONDITIONAL PROBABILITIES FOR THE LATENT CLASS INDICATORS

          PROBABILITY PROFILES
                Latent Class
                   1      2      3      4
PROPERTY
     Category 1  0.839  0.315  0.874  0.975
     Category 2  0.161  0.685  0.126  0.025
FIGHT
     Category 1  0.632  0.399  0.916  0.986
     Category 2  0.368  0.601  0.084  0.014
SHOPLIFT
     Category 1  0.785  0.211  0.622  0.940
     Category 2  0.215  0.789  0.378  0.060
```

Copy | Save to file | Close

Calculator Convenience Feature: Copying Sets of Values for Covariates

- Enter your values for the first set
- Click on the button above the OK/Cancel/Apply called “Copy values to new set”
  - That will copy all the values you just entered to a new set
- A second “Default” tab will open to the right of the original Default tab with the copied values that you can edit
  - For clarity, you can change the label of either ”Default” tab
- Click OK
- Probabilities for both sets will be shown, one below the other

Using MODEL CONSTRAINTS for Direct Effect Probability Calculations for the Property Indicator: Class 1 with Male = 0 vs 1, All Other X's=0, Except Age94=Sample Mean

```
MODEL: %OVERALL%
       c ON male-age94;
       property ON male (b1)
       black divorce
       age94 (b2);
       fight ON male hisp age94;
       shoplift ON black hisp;
       threat ON black;
       pot ON age94;
       drug ON black age94;
       goods ON male black;

       %c#1%
       [property$1] (t1);
```

```
MODEL CONSTRAINT:   NEW(logit0 logit1 prob0 prob1);
                    ! suffix of 0/1 corresponds
                    ! to male = 0/1
                    ! sample mean of age94 = 2.957
                    logit0 = -t1 + b2*2.957;
                    logit1 = -t1 + b1 + b2*2.957;
                    prob0 = 1/(1+EXP(-logit0));
                    prob1 = 1/(1+EXP(-logit1));
```

Outline

- Background
- Antisocial behavior example
- Searching for direct effects
    - Different approaches, Penalized SEM (PSEM)
    - ASB example (reduced version)
        - Latent classes and C ON X results
- Evaluating direct effects
    - Direct effects results, calculator
- **Class-varying direct effects**
    - PSEM DIFF priors, ALIGN option
- Multistep analysis with direct effects and distal outcomes
- Recap and further research

Input for PSEM DIFF Priors: Model 10, 15 Direct Effects

```
TITLE:      7U-7X, 15 direct effects of model 10
            PSEM (1.0) with class-varying direct
            effects and alignment output

DATA:       FILE = asbfree.dat;

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 fight
            shoplift threat pot drug goods sex black
            hisp single divorce dropout age94;

            CATEGORICAL = property-goods;
            CLASSES = c(4);

ANALYSIS:   TYPE = MIXTURE;
            ESTIMATOR - ML;
            STARTS = 400 100;
            PROCESSORS = 12;

MODEL:      %OVERALL%
            c ON sex-age94;

            property ON sex black divorce age94;
            fight ON sex hisp age94;
            shoplift ON black hisp;
            threat ON black;
            pot ON age94;
            drug ON black age94;
            goods ON sex black;

            %c#1%
            property ON sex black divorce age94 (a1-a4);
            fight ON sex hisp age94 (a5-a7);
            shoplift ON black hisp (a8-a9);
            threat ON black (a10);
            pot ON age94 (a11);
            drug ON black age94 (a12-a13);
            goods ON sex black (a14-a15);
```

Input for PSEM DIFF Priors, Continued

```
%c#2%
property ON sex black divorce age94 (b1-b4);
fight ON sex hisp age94 (b5-b7);
shoplift ON black hisp (b8-b9);
threat ON black (b10);
pot ON age94 (b11);
drug ON black age94 (b12-b13);
goods ON sex black (b14-b15);

%c#3%
property ON sex black divorce age94 (c1-c4);
fight ON sex hisp age94 (c5-c7);
shoplift ON black hisp (c8-c9);
threat ON black (c10);
pot ON age94 (c11);
drug ON black age94 (c12-c13);
goods ON sex black (c14-c15);

%c#4%
property ON sex black divorce age94 (d1-d4);
fight ON sex hisp age94 (d5-d7);
shoplift ON black hisp (d8-d9);
threat ON black (d10);
pot ON age94 (d11);
drug ON black age94 (d12-d13);
goods ON sex black (d14-d15);
```

MODEL
PRIOR: **DO(#,1,15) DIFF(a# b# c# d#)∼ALF(0,1.0)**;

OUTPUT: ALIGN;

Results for Class-Varying Direct Effects: Model 10, 15 Direct Effects

- Class-varying direct effects using PSEM 1.0 with DIFF priors: logL = -23,680 (BIC not available)
- Class-varying direct effects using non-PSEM: logL = -23,668, BIC = 48,333
  - 2 fixed direct effects (couldn't be estimated)
- Class-invariant direct effects using non-PSEM: BIC = 48,026
  - Better BIC than for class-varying direct effects
  - No need for class-varying direct effects in this example
- PSEM with DIFF priors useful with class-varying direct effects
  - The ALIGN output option shows which effects have significant differences across classes - see also factor analysis alignment in Asparouhov & Muthén (2014) and other papers at
    - `https://www.statmodel.com/MeasurementInvariance.shtml`

Alignment Output with PSEM DIFF Priors

- a1, b1, c1, d1: Property on sex (male) comparing the 4 classes

DIFF ANALYSIS FOR PARAMETERS

A1 B1 C1 D1

Chi-square value	0.053
Degrees of freedom	3
P-value	0.997

Param	Param	Value	Value	Difference	SE	P-value
B1	A1	0.723	0.563	0.160	0.750	0.831
C1	A1	0.723	0.563	0.159	0.742	0.830
C1	B1	0.723	0.723	0.000	0.022	0.985
D1	A1	0.724	0.563	0.160	0.754	0.831
D1	B1	0.724	0.723	0.001	0.021	0.980
D1	C1	0.724	0.723	0.001	0.023	0.967

Approximate Invariance Holds For:

A1 B1 C1 D1

Average Value Across Invariant Parameters: 0.683

Invariant Values, Difference to Average and Significance

Param	Value	Difference	SE	P-value
A1	0.563	-0.120	0.310	0.699
B1	0.723	0.040	0.458	0.931
C1	0.723	0.040	0.450	0.930
D1	0.724	0.040	0.461	0.930

Alignment Output, Continued

- a11, b11, c11, d11: Pot on age94 comparing the 4 classes

DIFF ANALYSIS FOR PARAMETERS
A11 B11 C11 D11

Chi-square value	7.795
Degrees of freedom	3
P-value	0.050

Param	Param	Value	Value	Difference	SE	P-value
B11	A11	0.502	0.240	0.261	0.174	0.133
C11	A11	-0.813	0.240	-1.053	0.513	0.040
C11	B11	-0.813	0.502	-1.315	0.555	0.018
D11	A11	0.176	0.240	-0.064	0.047	0.178
D11	B11	0.176	0.502	-0.325	0.174	0.062
D11	C11	0.176	-0.813	0.990	0.510	0.052

Approximate Invariance Was Not Found.

- c11 vs b11 shows the largest difference: Pot on age94, comparing the third and second classes - Pot and High

Outline

- Background
- Antisocial behavior example
- Searching for direct effects
    - Different approaches, Penalized SEM (PSEM)
    - ASB example (reduced version)
        - Latent classes and C ON X results
- Evaluating direct effects
    - Direct effects results, calculator
- Class-varying direct effects
    - PSEM DIFF priors, ALIGN option
- **Multistep analysis with direct effects and distal outcomes**
- Recap and further research

Multistep Analysis with Direct Effects and Distal Outcomes

- Adding distal outcomes to the analysis with direct effects
  - The first step measurement model includes the covariates and their direct effects
  - What should the last step look like?
    - 3-step
    - BCH
    - 2-step
- Example: ASB data with the 15 direct effects of Model 10, adding alcohol dependence and abuse as distal outcomes
  - Strongly skewed variables with large floor effects; treated as continuous in these analyses
  - Likely to be correlated even when conditioned on latent class

First Step, Combined Approach with Direct Effects

TITLE: First step, combined approach, measurement model including direct effects

DATA: FILE = asbfree.dat;

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;

USEVAR = property fight shoplift threat pot drug goods sex black hisp single divorce dropout age94;

AUXILIARY = dep abuse;

CATEGORICAL = property-goods;

CLASSES= c(4);

ANALYSIS: TYPE = MIXTURE;

ESTIMATOR = ML;

OPTSEED = 21345;

PROCESSORS = 12;

MODEL: %OVERALL%

c ON sex-age94;

! Direct effects:

property ON sex black divorce age94;

fight ON sex hisp age94;

shoplift ON black hisp;

threat ON black;

pot ON age94;

drug ON black age94;

goods ON sex black;

SAVEDATA: SAVE = cprob bchweights;

FILE = final.dat;

OUTPUT: ! Re-ordering the classes

! based on previous run:

SVALUES(1 4 2 3);

Input for 3-Step: Last Step

```
TITLE:
            Last step of 3-step

DATA:       FILE = final.dat;

VARIABLE:   NAMES = property fight shoplift threat
            pot drug goods sex black hisp single
            divorce dropout age94 dep abuse w1-w4
            cprob1-cprob4 n;

            USEVARIABLES = sex-age94
            dep abuse n;

            NOMINAL = n;

            CLASSES = c(4);

ANALYSIS:   TYPE = MIXTURE;
            ESTIMATOR = ML;
            STARTS = 0; ! STARTS=800 200 gives
            ! another solution
            PROCESSORS = 12;

MODEL:      %OVERALL%

            dep abuse ON sex-age94;
            dep WITH abuse;
            ! No direct effects since indicators
            ! not in the model

            %c#1%
            [n#1@1.616 n#2@-1.634 n#3@-1.108];
            %c#2%
            [n#1@4.971 n#2@6.760 n#3@4.382];
            %c#3%
            [n#1@-0.101 n#2@-0.734 n#3@1.908];
            %c#4%
            [n#1@-2.518 n#2@-9.861 n#3@-3.323];
```

Input for BCH: Last Step

```
TITLE:        Last step of BCH

DATA:         FILE =final.dat;

VARIABLE:     NAMES = property fight shoplift threat
              pot drug goods sex black hisp single
              divorce dropout age94 dep abuse w1-w4
              cprob1-cprob4 n;

              USEVARIABLES = dep abuse
              sex-age94 w1-w4;

              TRAINING = w1-w4(BCH);

              CLASSES = c(4);

ANALYSIS:     TYPE = MIXTURE;
              ESTIMATOR = ML;
              STARTS = 0;
              PROCESSORS = 12;

MODEL:        %OVERALL%
              dep abuse ON sex-age94;
              dep WITH abuse;
              ! No direct effects since indicators
              ! not in the model
```

Input for 2-Step: Last Step

```
TITLE:
            Last step of 2-step

DATA:       FILE IS asbfree.dat;

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 fight
            shoplift threat pot drug goods sex black
            hisp single divorce dropout age94 dep
            abuse;

            CATEGORICAL = property-goods;

            CLASSES = c(4);
```

```
ANALYSIS:   TYPE = MIXTURE;
            ESTIMATOR - ML;
            STARTS = 0; ! STARTS=400 100 gives
            another solution
            PROCESSORS = 12;

MODEL:      %OVERALL%

            dep abuse ON sex-age94;
            dep WITH abuse;
            ! Input continues on next slides
```

Input for 2-Step, Last Step, Continued

```
! Direct effects need to be included
! because the indicators are in the model

! Output from SVALUES.
! First for OVERALL,
! then class-specific
property ON sex;
property ON black;
property ON divorce;
property ON age94;
fight ON sex;
fight ON hisp;
fight ON age94;
shoplift ON black;
shoplift ON hisp;
threat ON black;
pot ON age94;
drug ON black;
drug ON age94;
goods ON sex;
goods ON black;
```

```
c#1 ON sex@1.28575;
c#1 ON black@1.26114;
c#1 ON hisp@-0.25585;
c#1 ON single@0.01131;
c#1 ON divorce@0.27248;
c#1 ON dropout@0.28195;
c#1 ON age94@-0.20425;
c#2 ON sex@1.91380;
c#2 ON black@0.52702;
c#2 ON hisp@-0.65689;
c#2 ON single@0.43056;
c#2 ON divorce@0.79212;
c#2 ON dropout@0.62244;
c#2 ON age94@-0.23210;
c#3 ON sex@-0.01762;
c#3 ON black@-1.07972;
c#3 ON hisp@-0.66859;
c#3 ON single@0.31222;
c#3 ON divorce@0.46191;
c#3 ON dropout@-0.06458;
c#3 ON age94@-0.03458;
```

```
[ c#1@-0.64931 ];
[ c#2@-1.86032 ];
[ c#3@-0.49162 ];

%C#1%

property ON sex@0.67567 (29);
property ON black@-0.74400 (30);
property ON divorce@-0.45823 (31);
property ON age94@-0.09999 (32);
fight ON sex@0.58659 (33);
fight ON hisp@0.73120 (34);
fight ON age94@-0.10968 (35);
shoplift ON black@-0.16065 (36);
shoplift ON hisp@0.45153 (37);
threat ON black@-0.71839 (38);
pot ON age94@0.22548 (39);
drug ON black@-0.81339 (40);
drug ON age94@0.22387 (41);
goods ON sex@0.83449 (42);
goods ON black@-0.54608 (43);

[ property$1@1.35209 ];
[ fight$1@0.21694 ];
[ shoplift$1@1.29552 ];
[ threat$1@-0.69313 ];
[ pot$1@1.35039 ];
[ drug$1@3.95855 ];
[ goods$1@2.79884 ];
```

%C#2%

```
property ON sex@0.67567 (29);
property ON black@-0.74400 (30);
property ON divorce@-0.45823 (31);
property ON age94@-0.09999 (32);
fight ON sex@0.58659 (33);
fight ON hisp@0.73120 (34);
fight ON age94@-0.10968 (35);
shoplift ON black@-0.16065 (36);
shoplift ON hisp@0.45153 (37);
threat ON black@-0.71839 (38);
pot ON age94@0.22548 (39);
drug ON black@-0.81339 (40);
drug ON age94@0.22387 (41);
goods ON sex@0.83449 (42);
goods ON black@-0.54608 (43);

[ property$1@-1.07329 ];
[ fight$1@-0.73433 ];
[ shoplift$1@-1.31987 ];
[ threat$1@-1.69122 ];
[ pot$1@-1.69565 ];
[ drug$1@0.13231 ];
[ goods$1@0.34741 ];

! Etc for classes 3 and 4
```

Comparing Results for 3-Step, BCH, and 2-Step

- Class probabilities from the first step (Model 10, 15 direct effects):
  - 0.321 (Threat/Fight), 0.130 (High), 0.172 (Drugs), 0.377 (low)
- Class probabilities in last step (STARTS=0):
  - 3-step: 0.461, 0.019, 0.164, 0.356
    - Failure - simulations in Asparouhov & Muthén (2014; Web Note 15) show undesirable influence of continuous distal outcomes
  - BCH: 0.321, 0.130, 0.173, 0.377
  - 2-step: 0.321, 0.130, 0.172, 0.377
- STARTS > 0
  - 3-step: Better LL but not replicated, class probabilities still different from first step, different estimates for distal outcomes regressed on covariates - solution influenced by distal outcomes
  - BCH: Same solution as for STARTS=0
  - 2-step: Better LL, same class probabilities, different estimates for distal outcomes regressed on covariates - solution influenced by distal outcomes
- BCH preferrable (see also Asparouhov & Muthén, Web Note 21 and Web Talk 8) due to within-class non-normality of distal outcomes

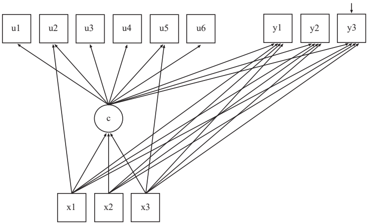
Distal Outcomes for BCH: Comparing Y ON X with Direct Effects in 1st Step vs No Covariates, No Directs in 1st Step

	1st step: Covariates and 15 direct effects (Model 10)			1st step: Latent class indicators only		
	Estimate	S.E.	Est./S.E.	Estimate	S.E.	Est./S.E.
DEP ON						
SEX	0.140	0.027	5.147	0.158	0.025	6.248
BLACK	-0.031	0.032	-0.979	0.001	0.028	0.018
HISP	-0.008	0.032	-0.255	-0.022	0.032	-0.702
SINGLE	0.250	0.029	8.585	0.251	0.029	8.647
DIVORCE	0.154	0.032	4.863	0.163	0.032	5.171
DROPOUT	0.362	0.036	10.171	0.374	0.035	10.558
AGE94	0.004	0.006	0.801	0.000	0.006	0.031
ABUSE ON						
SEX	0.268	0.032	8.386	0.274	0.030	9.264
BLACK	-0.294	0.037	-7.939	-0.269	0.033	-8.105
HISP	-0.051	0.038	**-1.363**	-0.094	0.037	**-2.529**
SINGLE	0.353	0.034	10.310	0.359	0.034	10.525
DIVORCE	0.257	0.037	6.913	0.277	0.037	7.493
DROPOUT	0.150	0.042	3.591	0.169	0.042	4.060
AGE94	-0.013	0.007	**-1.909**	-0.027	0.007	**-4.092**

Outline

- Background
- Antisocial behavior example
- Searching for direct effects
  - Different approaches, Penalized SEM (PSEM)
  - ASB example (reduced version)
    - Latent classes and C ON X results
- Evaluating direct effects
  - Direct effects results, calculator
- Multistep analysis with direct effects and distal outcomes
- **Recap and further research**

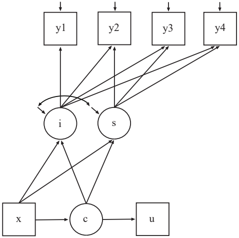
Recap of Alternative Approaches



- Analyses without the distal outcomes
  - (1) Including direct effects
  - (2) Ignoring direct effects, including X's
  - Approaches (1) and (2) were compared with respect to C ON X
- Analyses with the distal outcomes: Multistep such as BCH
  - (3) Including direct effects in first step
  - (4) Ignoring direct effects in first step
  - (5) Including direct effects only in last step (for 2-step)
  - (6) Excluding covariates in first step (measurement model using indicators only), adding them in last step, ignoring direct effects
  - Approaches (3) and (6) were compared with respect to Y ON X

Further Research

- Monte Carlo simulation studies
  - PSEM approach
  - Figure 12 of Asparouhov & Muthén (2025)
  - MplusAutomation `https://www.statmodel.com/usingmplusviar.shtml`
- PSEM needs work for the ordinal case (avoiding threshold collapse)
- Other measurement models:
  - Growth mixture modeling:
    - Muthén & Shedden (1999), Muthén et al. (2002), Muthén (2004), Asparouhov & Muthén (2014; Web Note 15)



References

- Asparouhov & Muthén (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. Structural Equation Modeling: A Multidisciplinary Journal, 21:3, 329-341. The posted version corrects several typos in the published version. An earlier version of this paper was posted as Mplus Web Notes: No. 15.
- Asparouhov & Muthén (2021). Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. Mplus Web Notes: No 21.
- Asparouhov & Muthén (2024). Penalized structural equation models. Structural Equation Modeling: A Multidisciplinary Journal, 31(3), 429–454.
- Asparouhov & Muthén (2025). Methodological advances with penalized structural equation models. Structural Equation Modeling: A Multidisciplinary Journal, 32(4), 688-716.
- Muthén & Shedden (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics, 55, 463-469.
- Muthén et al. (2002). General growth mixture modeling for randomized preventive interventions. Biostatistics, 3 (4), 459-475.
- Muthén (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (ed.), Handbook of quantitative methodology for the social sciences (pp. 345-368). Newbury Park, CA: Sage Publications.
  `https://www.statmodel.com/download/KaplanChapter19.pdf`
- Muthén (2025). Mplus: An overview of its unique analysis capabilities. Forthcoming in the Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences: Volume Three.