# Using Mplus To Do Multistep Mixture Modeling: Latent Transition Analysis

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

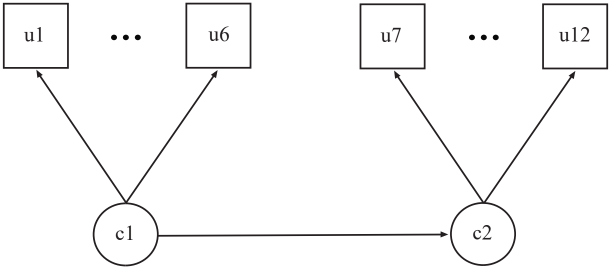
Tihomir Asparouhov
Mplus

Mplus Web Talks: No. 10
August 2025

We thank Thuy Nguyen and Noah Hastings for expert assistance.

# Outline

# Regular LTA (Binary Indicators)




1. Indicator probabilities: $P(U_{jt}|C_t)$ - LCA for each timepoint
2. Initial status probabilities: $P(C_1)$
3. Transition probabilities: $P(C_2|C_1)$

- Indicators can be continuous, binary, ordinal, nominal, counts
- For a discussion of regular LTA, see Mplus Web Talk 2

# Input for Regular LTA

| | |
|---|---|
| TITLE: | LTA using binary u1-u12 latent class indicators |
| DATA: | FILE = LTADistalN2K.dat;<br>! Simulated data, N = 2,000<br>! Entropy: C1 = 0.646, C2 = 0.656 |
| VARIABLE: | NAMES = u1-u12 x1 y1 y2 x2 y3 x3 class1 class2;<br>USEVARIABLES = u1-u12;<br>MISSING = ALL(999);<br>CLASSES = c1(3) c2(3);<br>CATEGORICAL = u1-u12; |
| ANALYSIS: | TYPE = MIXTURE;<br>STARTS = 40 10; |

# Input Regular LTA, Continued

```
                                    ! In a typical LTA application,
                                    ! u7-u12 are the same measures
MODEL:     %OVERALL%                ! as u1-u6 but recorded at time 2.
           c2 ON c1;                ! To ensure capturing the same
           ! This produces estimates for   ! latent class construct at the
           ! C1#1, C1#2, C2#1, C2#2        ! two timepoints, measurement
           ! C2#1 ON C1#1, C2#2 ON C1#1    ! invariance across time is specified
           ! C2#1 ON C1#2, C2#2 ON C1#2    ! by equality constraints:

MODEL C1:  %C1#1%                   MODEL C2:  %C2#1%
           [u1$1-u6$1] (1-6);                  [u7$1-u12$1] (1-6);

           %C1#2%                              %C2#2%
           [u1$1-u6$1] (11-16);                [u7$1-u12$1] (11-16);

           %C1#3%                              %C2#3%
           [u1$1-u6$1] (21-26);                [u7$1-u12$1] (21-26);
```

MEAN/PROBABILITY PROFILES FOR C1 (same for C2)

| | Latent class | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| U1 | | | |
| Category 1 | 0.206 | 0.858 | 0.798 |
| Category 2 | 0.794 | 0.142 | 0.202 |
| U2 | | | |
| Category 1 | 0.214 | 0.778 | 0.828 |
| Category 2 | 0.786 | 0.222 | 0.172 |
| U3 | | | |
| Category 1 | 0.232 | 0.786 | 0.801 |
| Category 2 | 0.768 | 0.214 | 0.199 |
| U4 | | | |
| Category 1 | 0.217 | 0.159 | 0.796 |
| Category 2 | 0.783 | 0.841 | 0.204 |
| U5 | | | |
| Category 1 | 0.211 | 0.200 | 0.781 |
| Category 2 | 0.789 | 0.800 | 0.219 |
| U6 | | | |
| Category 1 | 0.217 | 0.178 | 0.795 |
| Category 2 | 0.783 | 0.822 | 0.205 |

# Latent Transition Results

- Latent class probabilities
  - C1: 0.279, 0.299, 0.422. C2: 0.284, 0.263, 0.453

LATENT TRANSITION PROBABILITIES BASED ON THE ESTIMATED MODEL
C1 Classes (Rows) by C2 Classes (Columns)

| | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0.483 | 0.229 | 0.288 |
| 2 | 0.253 | 0.318 | 0.428 |
| 3 | 0.174 | 0.246 | 0.580 |

TRANSITION PROBABILITY ODDS

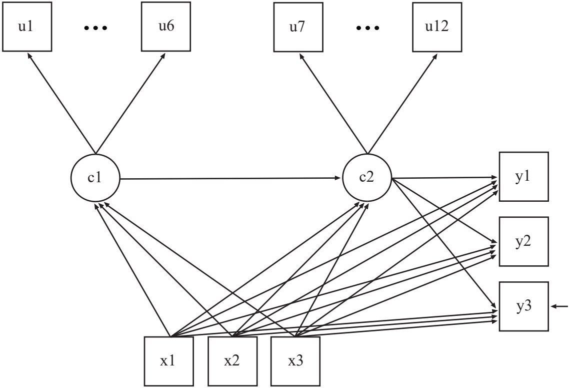
TRANSITION TABLE ODDS AND 95% CONFIDENCE INTERVALS FOR C1 TO C2

| | | |
|---|---|---|
| 1.000(1.000,1.000) | 0.474(0.327,0.687) | 0.596(0.458,0.775) |
| 0.795(0.551,1.148) | 1.000(1.000,1.000) | 1.344(0.944,1.914) |
| 0.299(0.224,0.399) | 0.424(0.316,0.569) | 1.000(1.000,1.000) |

# Outline

- Regular LTA
- **Multistep LTA with distal outcomes and covariates with missing data**
  - 3-step, BCH, and 2-step using imputation
  - Inputs, results
- LTA with transition-specific distal outcome means, dot language
- Empty cells
  - BCH failure
- Multiple-group LTA
- RI-LTA
- LCA at two timepoints

# LTA with Covariates and Distal Outcomes




- Covariates: x1 binary, x2-x3 continuous. Missing data on x1-x2
- Distal outcomes: y1 binary, y2 ordinal (3 categories), y3 continuous
- No interaction between C1 and X's in their influence on C2
- Multistep idea: Covariates and distal outcomes should not affect the measurement part of the model

# Multistep Rationale

- There is a need to separate the estimation of the measurement model from the full model so the latent classes don't change
  - Distal outcomes in the full model may be measured later in time than the variables of the measurement model
  - Covariates in the full model may affect the latent class enumeration
- The separation of the estimation is done by multistep approaches
  - First step: The latent classes are determined only by the measurement model
  - Last step: Distal outcomes and/or covariates of the full model are added
  - Theory is shown in the Statistical Methodology section of Mplus Web Talk 8

# Multistep Approaches: 3-Step, BCH, 2-Step

- The 3-step and BCH approaches use the posterior probabilities from the estimated measurement model to determine each individual's most likely class membership and take its measurement error into account
  - 3-step: Last step uses fixed logits from measurement model to take measurement error into account in the Most Likely Class (MLC) classification
  - BCH: Last step uses weights from measurement model to take measurement error into account in the MLC classification
- 2-step: Last step uses fixed measurement model parameters to avoid measurement error

# Studies of Multistep Approaches for LTA

- Simulations in Asparouhov & Muthén (Web Note 15 and 21):
  - Covariates
    - 3-step (WN15)
    - BCH (WN21)
  - Distal outcomes
    - BCH (WN21)
    - BCH does better than 2-step when distals are non-normal within class
  - 2-step harder to study because for each replication, the parameter values need to be changed
- Special case: Missing data on covariates used in the last step analysis

# Multistep Approaches with Missing Data on X In the Last Step Analysis

- Covariates brought into the model:
  - ML and numerical integration
    - 3-step, BCH not available with numerical integration, 2-step
    - Pros: Likelihood-ratio testing available
    - Cons: Cannot handle many covariates with missing, can be somewhat unstable with large amounts of missing data, can be imprecise without many integration points, cannot specify categorical covariates
  - Bayes
    - 3-step, BCH not available due to weights, 2-step
    - Pros: Stable
    - Cons: No LRT, can be slow, cannot specify categorical covariates
- **Multiple imputation**:
  - 3-step, BCH, 2-step
  - Pros: Stable, can specify categorical covariates, can use non-model variables for missing data imputation
  - Cons: No LRT, slow with many covariates with missing
- Corresponding LCA approaches are discussed in Mplus Web Talk 8

# Multistep LTA: BCH Weight Computation in the First Step

- Web Talk 8 discusses BCH weights for LCA (one latent class variable)
- Mplus obtains BCH weights for LTA as the product of weights for each timepoint estimated separately
- In a joint analysis of all timepoints, independence of timepoints is obtained in the estimation by not regressing C on C, so that the weights for each timepoint can be automatically multiplied together to create weights for all timepoints
    - First step: Don't include C on C
    - Last step: Include C on C
        - See simulation studies in Asparouhov & Muthén (Web Note 21)
        - This agrees with Bakk, Tekle, Vermunt (2013; p. 288, p. 303)

# Outline

- Regular LTA
- Multistep LTA with distal outcomes and covariates with missing data
  - 3-step, BCH, and 2-step using imputation
  - **Inputs, results**
- LTA with transition-specific distal outcome means, dot language
- Empty cells
  - BCH failure
- Multiple-group LTA
- RI-LTA
- LCA at two timepoints

# Analysis Stages for Combined Approach

Stage A. Measurement Model Analysis

| Input Specifications | Variables Saved (SAVEDATA*) |
|---|---|
| USEV = u1-u12<br>AUXILIARY (for final stage)<br>SAVEDATA CPROB (for 3-step)<br>SAVEDATA BCHWEIGHTS (for BCH)<br>SVALUES (for 2-step) | u1-u12*<br>x1-x3, y1-y3<br>cprob1-cprob9, n1, n2*<br>w1-w9* |

Stage B. Final Analysis (Multiple imputation + Full model analysis)

DATA IMPUTATION for missing on X (can be replaced by integration)

| 3-Step | BCH | 2-Step |
|---|---|---|
| USEV = n1, n2, x1-x3, y1-y3<br>Logits from Stage A<br>(manually added) | USEV = x1-x3, y1-y3<br>TRAINING = w1-w9(BCH) | USEV = u1-u12,x1-x3,y1-y3<br>SVALUES from Stage A<br>(manually added) |

# Stage A: Saving Information for Final Analysis

```
TITLE:        Stage A

DATA:         FILE = LTADistalN2K.dat;
              ! Same dataset as used for the regular LTA analysis

VARIABLE:     NAMES = u1-u12 x1 y1 y2 x2 y3 x3 class1 class2;
              USEVARIABLES = u1-u12;
              MISSING = ALL(999);
              CLASSES = c1(3) c2(3);
              CATEGORICAL = u1-u12;
              AUXILIARY = x1 x2 x3 y1 y2 y3;
              ! Saving the covariates and distals for the last step

ANALYSIS:     TYPE = MIXTURE;
              STARTS = 100 40;
              PROCESSORS = 8;

              ! Input continues on the next slide
```

# Stage A, Continued

```
MODEL:        %OVERALL%
              ! Note: no c2 ON c1 in the first step
              ! accomodates all 3 multistep methods

MODEL C1:     %C1#1%
              [u1$1-u6$1] (1-6);

              %C1#2%
              [u1$1-u6$1] (11-16);

              %C1#3%
              [u1$1-u6$1] (21-26);

MODEL C2:     %C2#1%
              [u7$1-u12$1] (1-6);

              %C2#2%
              [u7$1-u12$1] (11-16);

              %C2#3%
              [u7$1-u12$1] (21-26);

SAVEDATA:     SAVE = CPROB;
              SAVE = BCHWEIGHTS;
              FILE = stageA.dat;

OUTPUT:       SVALUES;
```

FINAL CLASS COUNTS AND PROPORTIONS FOR EACH LATENT CLASS VARIABLE BASED ON THE ESTIMATED MODEL

| Latent Class Variable | Class | | |
|---|---|---|---|
| C1 | 1 | 848.34955 | 0.42417 |
| | 2 | 554.73920 | 0.27737 |
| | 3 | 596.91125 | 0.29846 |
| C2 | 1 | 906.45416 | 0.45323 |
| | 2 | 568.04913 | 0.28402 |
| | 3 | 525.49677 | 0.26275 |

- Reordering to increasing class size based on C1:
  - SVALUES (2 3 1 | 2 3 1);

# Stage A: Reordering of the Classes

```
TITLE:          Stage A

DATA:           FILE = LTADistalN2K.dat;

VARIABLE:       NAMES = u1-u12 x1 y1 y2 x2 y3 x3 class1 class2;
                USEVARIABLES = u1-u12;
                MISSING = ALL(999);
                CLASSES = c1(3) c2(3);
                CATEGORICAL = u1-u12;
                AUXILIARY = x1 x2 x3 y1 y2 y3;

ANALYSIS:       TYPE = MIXTURE;
                STARTS = 0;
                OPTSEED = 915642;
                PROCESSORS = 8;

                ! Input continues on the next slide
```

# Stage A Reorder, Continued

```
MODEL:        %OVERALL%

MODEL C1:     %C1#1%
              [u1$1-u6$1] (1-6);

              %C1#2%
              [u1$1-u6$1] (11-16);

              %C1#3%
              [u1$1-u6$1] (21-26);

MODEL C2:     %C2#1%
              [u7$1-u12$1] (1-6);

              %C2#2%
              [u7$1-u12$1] (11-16);

              %C2#3%
              [u7$1-u12$1] (21-26);

SAVEDATA:     FILE = stageA.dat;
              SAVE = BCHWEIGHTS;
              SAVE = CPROB;

OUTPUT:       SVALUES (2 3 1 | 2 3 1);
```

# Latent Class Indicator Probabilities

| Regular LTA ( C on C) | | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| U1 | | | |
| Category 1 | 0.206 | 0.858 | 0.798 |
| Category 2 | 0.794 | 0.142 | 0.202 |
| U2 | | | |
| Category 1 | 0.214 | 0.778 | 0.828 |
| Category 2 | 0.786 | 0.222 | 0.172 |
| U3 | | | |
| Category 1 | 0.232 | 0.786 | 0.801 |
| Category 2 | 0.768 | 0.214 | 0.199 |
| U4 | | | |
| Category 1 | 0.217 | 0.159 | 0.796 |
| Category 2 | 0.783 | 0.841 | 0.204 |
| U5 | | | |
| Category 1 | 0.211 | 0.200 | 0.781 |
| Category 2 | 0.789 | 0.800 | 0.219 |
| U6 | | | |
| Category 1 | 0.217 | 0.178 | 0.795 |
| Category 2 | 0.783 | 0.822 | 0.205 |

| First Step (no C on C) | | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| U1 | | | |
| Category 1 | 0.207 | 0.854 | 0.799 |
| Category 2 | 0.793 | 0.146 | 0.201 |
| U2 | | | |
| Category 1 | 0.213 | 0.776 | 0.829 |
| Category 2 | 0.787 | 0.224 | 0.171 |
| U3 | | | |
| Category 1 | 0.229 | 0.789 | 0.800 |
| Category 2 | 0.771 | 0.211 | 0.200 |
| U4 | | | |
| Category 1 | 0.219 | 0.159 | 0.793 |
| Category 2 | 0.781 | 0.841 | 0.207 |
| U5 | | | |
| Category 1 | 0.213 | 0.199 | 0.779 |
| Category 2 | 0.787 | 0.801 | 0.221 |
| U6 | | | |
| Category 1 | 0.216 | 0.173 | 0.797 |
| Category 2 | 0.784 | 0.827 | 0.203 |

# Stage B: Last Step of 3-Step (Based on the Reordering)

| | |
|---|---|
| TITLE: | Imputing missing on X and doing last step analysis of the full model using 3-step |
| DATA: | **FILE = stageA.dat;** |
| VARIABLE: | **NAMES = u1-u12 x1-x3 y1-y3 w1-w9 cprob1-cprob9 n1 n2 mlcjoint**; ! Variables listed in stage A output<br>**USEVARIABLES = n1 n2 x1-x3 y1-y3;**<br>CATEGORICAL = y1 y2;<br>NOMINAL = n1 n2;<br>MISSING = *;<br>CLASSES = c1(3) c2(3); |
| **DATA IMPUTATION:** | **NAMES = u1-u12(C) x1(C) x2 x3 y1(C) y2(C) y3;**<br>! NAMES: variables that inform the imputation<br>**IMPUTE = x1(C) x2;**<br>**NDATASETS = 100;**<br>**SAVE = imputed*.dat;**<br>**THIN = 100;** |
| ANALYSIS: | TYPE = MIXTURE;<br>ESTIMATOR = MLR;<br>STARTS = 0;<br>! Input continues on the next slide |

# Last Step: 3-Step, Continued

```
MODEL:      %OVERALL%
            c2 ON c1;
            c1-c2 ON x1-x3;
            y1-y3 ON x1-x3;
            ! No missing on x after imputation,
            ! so regular covariates

MODEL C1:   %C1#1%
            [n1#1@2.607 n1#2@0.558];

            %C1#2%
            [n1#1@0.042 n1#2@2.294];

            %C1#3%
            [n1#1@-4.165 n1#2@-2.120];

Model C2:   %C2#1%
            [n2#1@2.575 n2#2@0.497];
            [y1$1 y2$1 y2$2 y3];

            %C2#2%
            [n2#1@0.155 n2#2@2.308];
            [y1$1 y2$1 y2$2 y3];

            %C2#3%
            [n2#1@-4.205 n2#2@-2.172];
            [y1$1 y2$1 y2$2 y3];

OUTPUT:     TECH8;
            ! Showing Bayes imputation iterations
```

- Logit values for each latent class variable found in the Stage A output

# Stage B: Last Step of BCH (Based on the Reordering)

```
TITLE:            Imputing missing on X and doing last step
                  analysis using BCH

DATA:             FILE = stageA.dat;

VARIABLE:         NAMES = u1-u12 x1-x3 y1-y3 w1-w9
                  cp1-cp9 n1 n2 mlcjoint;
                  USEVARIABLES = x1-x3 y1-y3;
                  CATEGORICAL = y1 y2;
                  MISSING = *;
                  CLASSES = c1(3) c2(3);
                  TRAINING = w1-w9(BCH);

DATA IMPUTATION:  NAMES = u1-u12(C) x1(C) x2 x3
                  y1(C) y2(C) y3;
                  IMPUTE = x1(C) x2;
                  NDATASETS = 100;
                  SAVE = imputed*.dat; ! Not needed
                  THIN = 100;

ANALYSIS:         TYPE = MIXTURE;
                  ESTIMATOR = MLR;
                  STARTS = 0;

MODEL:            %OVERALL%
                  c2 ON c1;
                  c1-c2 ON x1-x3;
                  y1-y3 ON x1-x3;

MODEL C2:         %C2#1%
                  [y1$1 y2$1 y2$2 y3];

                  %C2#2%
                  [y1$1 y2$1 y2$2 y3];

                  %C2#3%
                  [y1$1 y2$1 y2$2 y3];

OUTPUT:           TECH8;
```

# Stage B: Last step of 2-step (Based on the Reordering)

```
TITLE:              Imputing missing on X and doing last step analysis of
                    the full model using 2-step

DATA:               FILE = stageA.dat;
                    ! 2-step can alternatively use the original data
                    ! since most likely class and weights are not needed

VARIABLES:          NAMES = u1-u12 x1-x3 y1-y3 w1-w9 cp1-cp9 n1 n2
                    mlcjoint;
                    USEVARIABLES = u1-u12 x1-x3 y1-y3;
                    CATEGORICAL = u1-u12 y1 y2;
                    MISSING = *;
                    CLASSES = c1(3) c2(3);

DATA IMPUTATION:    NAMES = u1-u12(C) x1(C) x2 x3 y1(C) y2(C) y3;
                    IMPUTE = x1(C) x2;
                    NDATASETS = 100;
                    SAVE = imputed*.dat; ! Not needed
                    THIN = 100;

ANALYSIS:           TYPE = MIXTURE;
                    ESTIMATOR = MLR;
                    STARTS = 0;
```

# Last Step of 2-step, Continued

```
MODEL:      %OVERALL%
            c2 ON c1;
            c1-c2 ON x1-x3;
            y1-y3 ON x1-x3;

MODEL C1:   %C1#1%
            [ u1$1@-1.34037 ] (1);
            [ u2$1@-1.30798 ] (2);
            [ u3$1@-1.21206 ] (3);
            [ u4$1@-1.27215 ] (4);
            [ u5$1@-1.30615 ] (5);
            [ u6$1@-1.29090 ] (6);

            %C1#2%
            [ u1$1@1.76362 ] (7);
            [ u2$1@1.24352 ] (8);
            [ u3$1@1.31695 ] (9);
            [ u4$1@-1.66400 ] (10);
            [ u5$1@-1.39186 ] (11);
            [ u6$1@-1.56221 ] (12);

            %C1#3%
            [ u1$1@1.37946 ] (13);
            [ u2$1@1.57905 ] (14);
            [ u3$1@1.38788 ] (15);
            [ u4$1@1.34296 ] (16);
            [ u5$1@1.25728 ] (17);
            [ u6$1@1.37008 ] (18);
```

# LTA Stage B: 2-step, Continued

```
MODEL C2:   %C2#1%
            [ u7$1@-1.34037 ] (1);
            [ u8$1@-1.30798 ] (2);
            [ u9$1@-1.21206 ] (3);
            [ u10$1@-1.27215 ] (4);
            [ u11$1@-1.30615 ] (5);
            [ u12$1@-1.29090 ] (6);
            [y1$1 y2$1 y2$2 y3];

            %C2#2%
            [ u7$1@1.76362 ] (7);
            [ u8$1@1.24352 ] (8);
            [ u9$1@1.31695 ] (9);
            [ u10$1@-1.66400 ] (10);
            [ u11$1@-1.39186 ] (11);
            [ u12$1@-1.56221 ] (12);
            [y1$1 y2$1 y2$2 y3];

            %C2#3%
            [ u7$1@1.37946 ] (13);
            [ u8$1@1.57905 ] (14);
            [ u9$1@1.38788 ] (15);
            [ u10$1@1.34296 ] (16);
            [ u11$1@1.25728 ] (17);
            [ u12$1@1.37008 ] (18);
            [y1$1 y2$1 y2$2 y3];

OUTPUT:     TECH8;
```

# Results:
# LTA Class Probabilities

| Analysis | C1 | C2 |
|---|---|---|
| Regular LTA | 0.279, 0.299, 0.422 | 0.284, 0.263, 0.453 |
| Stage A | 0.277, 0.298, 0.424 | 0.284, 0.263, 0.453 |
| Stage B 3-step | 0.277, 0.298, 0.425 | 0.280, 0.268, 0.452 |
| Stage B BCH | 0.277, 0.299, 0.424 | 0.284, 0.263, 0.453 |
| Stage B 2-step | 0.278, 0.298, 0.424 | 0.277, 0.268, 0.455 |

# LTA BCH

Class probabilities: C1 = 0.277, 0.299, 0.424. C2 = 0.284, 0.263, 0.453

| | $X \rightarrow C2$ | | $X \rightarrow Y$ | |
|---|---|---|---|---|
| Covariate | C2#1 OR's | C2#2 OR's | Y1 OR's | Y3 Slopes |
| X1 | 0.246* | 0.523* | 1.885* | 0.424* |
| X2 | 1.641* | 1.271* | 1.578* | 0.499* |
| X3 | 0.699* | 0.815 | 1.502* | 0.580* |

| $C1 \rightarrow C2$ | | | $C2 \rightarrow Y$ | |
|---|---|---|---|---|
| Latent transition probability odds | | | Y1 ORs | Y3 Intercepts |
| 1.000 | 0.550* | 0.802 | 3.442 (1-3)* | C2#1 -0.939* |
| 0.698 | 1.000 | 1.144 | 1.135 (2-3) | C2#2 0.009 |
| 0.309* | 0.460* | 1.000 | | C2#3 1.092* |

Class probabilities: C1 = 0.278, 0.298, 0.424. C2 = 0.277, 0.268, 0.455

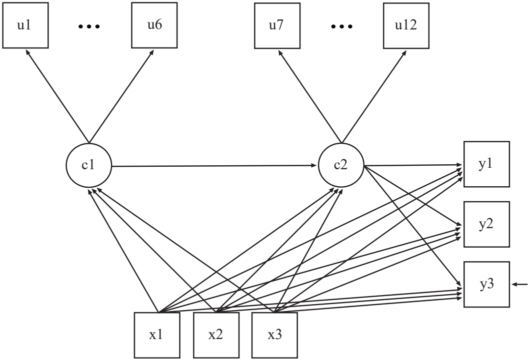
| | $X \rightarrow C2$ | | $X \rightarrow Y$ | |
|---|---|---|---|---|
| Covariate | C2#1 OR's | C2#2 OR's | Y1 OR's | Y3 Slopes |
| X1 | 0.266* | 0.504* | 1.806* | 0.471* |
| X2 | 1.692* | 1.301* | 1.580* | 0.499* |
| X3 | 0.695* | 0.813* | 1.506* | 0.580* |

| $C1 \rightarrow C2$ | | | $C2 \rightarrow Y$ | |
|---|---|---|---|---|
| Latent transition probability odds | | | Y1 ORs | Y3 Intercepts |
| 1.000 | 0.621* | 0.863 | 3.351 (1-3)* | C2#1 -0.952* |
| 0.700* | 1.000 | 1.234 | 1.029 (2-3) | C2#2 0.031 |
| 0.330* | 0.488* | 1.000 | | C2#3 1.015* |

# Outline

- Regular LTA
- Multistep LTA with distal outcomes and covariates with missing data
  - 3-step, BCH, and 2-step using imputation
  - Inputs, results
- **LTA with transition-specific distal outcome means, dot language**
- Empty cells
  - BCH failure
- Multiple-group LTA
- RI-LTA
- LCA at two timepoints

# LTA Model Specification Recap




- The figure specifies $C2 \rightarrow Y$: Y parameters mentioned in MODEL C2 classes lets them vary across C2 classes
  - This implies that Y parameters are held equal across C1 classes
- Y parameters **not** mentioned in MODEL C1 or MODEL C2 classes:
  - This implies that Y means/thresholds vary across all combinations of C1, C2 classes by default

# The Dot Language: Distal Mean/Threshold Specific to One LTA Transition

- Y parameters mentioned in MODEL C2 **and** in a specific class combination, that is, a certain transition pattern, using the dot language:
  - Example: % c1#2.c2#3%
- The dot specification overrides the Y parameter equalities across C1 classes implied by MODEL C2
- The dot specification can be placed in the %OVERALL% subsection of the MODEL command or below MODEL C2 under a new MODEL command section

# Last Step of 2-Step Based on Imputations Using the Dot Language

```
TITLE:              Imputing missing on X and doing last step analysis
                    of the full model using 2-step, adding one transition
                    using the dot laguage

DATA:               FILE = stageA.dat;

VARIABLES:          NAMES = u1-u12 x1-x3 y1-y3 w1-w9 cp1-cp9 n1 n2
                    mlcjoint;
                    USEVARIABLES = u1-u12 x1-x3 y1-y3;
                    CATEGORICAL = u1-u12 y1 y2;
                    MISSING = *;
                    CLASSES = c1(3) c2(3);

DATA IMPUTATION:    NAMES = u1-u12(C) x1(C) x2 x3 y1(C) y2(C) y3;
                    IMPUTE = x1(C) x2;
                    NDATASETS = 100;
                    SAVE = imputed*.dat; ! Not needed
                    THIN = 100;

ANALYSIS:           TYPE = MIXTURE;
                    ESTIMATOR = MLR;
                    STARTS = 0;
```

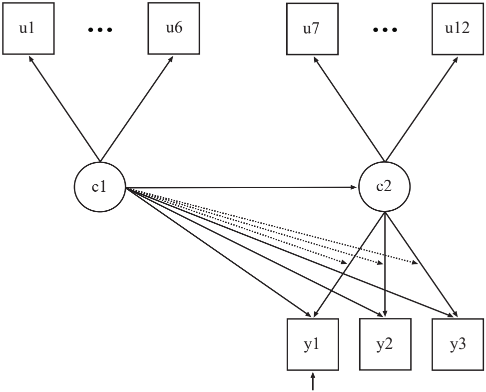
# LTA Stage B: 2-step, Continued

```
MODEL:      %OVERALL%
            c2 ON c1;
            c1-c2 ON x1-x3;
            y1-y3 ON x1-x3;

            ! Let the transition from class 2
            ! to class 3 have specific distal
            ! mean and thresholds:
            %C1#2. C2#3%
            [y1$1 y2$1 y2$2 y3];

MODEL C1:   %C1#1%
            [ u1$1@-1.34037 ] (1);
            [ u2$1@-1.30798 ] (2);
            [ u3$1@-1.21206 ] (3);
            [ u4$1@-1.27215 ] (4);
            [ u5$1@-1.30615 ] (5);
            [ u6$1@-1.29090 ] (6);

            %C1#2%
            [ u1$1@1.76362 ] (7);
            [ u2$1@1.24352 ] (8);
            [ u3$1@1.31695 ] (9);
            [ u4$1@-1.66400 ] (10);
            [ u5$1@-1.39186 ] (11);
            [ u6$1@-1.56221 ] (12);

            %C1#3%
            [ u1$1@1.37946 ] (13);
            [ u2$1@1.57905 ] (14);
            [ u3$1@1.38788 ] (15);
            [ u4$1@1.34296 ] (16);
            [ u5$1@1.25728 ] (17);
            [ u6$1@1.37008 ] (18);
```

# LTA Stage B: 2-step, Continued

```
MODEL C2:    %C2#1%
             [ u7$1@-1.34037 ] (1);
             [ u8$1@-1.30798 ] (2);
             [ u9$1@-1.21206 ] (3);
             [ u10$1@-1.27215 ] (4);
             [ u11$1@-1.30615 ] (5);
             [ u12$1@-1.29090 ] (6);
             [y1$1 y2$1 y2$2 y3];

             %C2#2%
             [ u7$1@1.76362 ] (7);
             [ u8$1@1.24352 ] (8);
             [ u9$1@1.31695 ] (9);
             [ u10$1@-1.66400 ] (10);
             [ u11$1@-1.39186 ] (11);
             [ u12$1@-1.56221 ] (12);
             [y1$1 y2$1 y2$2 y3];

             %C2#3%
             [ u7$1@1.37946 ] (13);
             [ u8$1@1.57905 ] (14);
             [ u9$1@1.38788 ] (15);
             [ u10$1@1.34296 ] (16);
             [ u11$1@1.25728 ] (17);
             [ u12$1@1.37008 ] (18);
             [y1$1 y2$1 y2$2 y3];

OUTPUT:      TECH8;
```

# Distal Means/Thresholds Specific to all LTA Transitions




- Distal outcomes: y1 continuous, y2 binary, y3 ordinal

# First Step of 3-Step

```
TITLE :        First step of 3-step. Using only u1-u12

DATA:          FILE = LTAdistalN2KDot.dat;

VARIABLE:      NAMES = u1-u12 x1 y2 y3 x2 y1 y3
               class1 class2;
               USEVARIABLES = u1-u12;
               MISSING = ALL(999);
               CLASSES = c1(3) c2(3);
               CATEGORICAL = u1-u12;
               AUXILIARY = x1 x2 x3 y1 y2 y3;

ANALYSIS:      TYPE = MIXTURE;
               OPTSEED = 533738;

MODEL:         %OVERALL%
               ! Note: No c2 ON c1 in the first step

MODEL C1:      %C1#1%
               [u1$1-u6$1] (1-6);

               %C1#2%
               [u1$1-u6$1] (11-16);

               %C1#3%
               [u1$1-u6$1] (21-26);

MODEL C2:      %C2#1%
               [u7$1-u12$1] (1-6);

               %C2#2%
               [u7$1-u12$1] (11-16);

               %C2#3%
               [u7$1-u12$1] (21-26);

SAVEDATA:      FILE = cprobdot.dat;
               SAVE = CPROB;

OUTPUT:        SVALUES(1 3 2 | 1 3 2);
```

# Last Step of 3-Step Using the Dot Language

```
TITLE:      Last step of 3-step.

DATA:       FILE = cprobdot.dat;

VARIABLE:   NAMES = u1-u12 x1-x3 y1-y3 cprob1-cprob9
            n1 n2 mlcjoint;
            USEVARIABLES = n1 n2 y1-y3;
            MISSING = *;
            CLASSES = c1(3) c2(3);
            CATEGORICAL = y2 y3;
            NOMINAL = n1 n2;

ANALYSIS:   TYPE = MIXTURE;
            STARTS= 0;

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

```
! Dot language specifications
! (labels not necessary):
%c1#1.c2#1%
[y1 y2$1 y3$1 y3$2] (m11-m14);
%c1#1.c2#2%
[y1 y2$1 y3$1 y3$2] (m21-m24);
%c1#1.c2#3%
[y1 y2$1 y3$1 y3$2] (m31-m34);

%c1#2.c2#1%
[y1 y2$1 y3$1 y3$2] (m41-m44);
%c1#2.c2#2%
[y1 y2$1 y3$1 y3$2] (m51-m54);
%c1#2.c2#3%
[y1 y2$1 y3$1 y3$2] (m61-m64);

%c1#3.c2#1%
[y1 y2$1 y3$1 y3$2] (m71-m74);
%c1#3.c2#2%
[y1 y2$1 y3$1 y3$2] (m81-m84);
%c1#3.c2#3%
[y1 y2$1 y3$1 y3$2] (m91-m94);
```

# Last Step of 3-Step Using the Dot Language, Continued

```
Model C1:
            %C1#1%
            [n1#1@2.757 n1#2@0.652];

            %C1#2%
            [n1#1@-0.089 n1#2@1.960];

            %C1#3%
            [n1#1@-4.091 n1#2@-2.256];

Model C2:
            %C2#1%
            [n2#1@2.632 n2#2@0.499];

            %C2#2%
            [n2#1@-0.173 n2#2@1.985];

            %C2#3%
            [n2#1@-4.128 n2#2@-2.346];

OUTPUT:     TECH15;
```

# Outline

- Regular LTA
- Multistep LTA with distal outcomes and covariates with missing data
  - 3-step, BCH, and 2-step using imputation
  - Inputs, results
- LTA with transition-specific distal outcome means, dot language
- **Empty cells**
  - BCH failure
- Multiple-group LTA
- RI-LTA
- LCA at two timepoints

# Empty Cells

- LTA with empty cells for the latent classes, N = 2000
- C1(4), C2(4)
- Distal outcome means estimated as influenced by C2

# LTA Without Covariates and Distal Outcomes

```
TITLE:

DATA:        FILE = LTAEmptyCellsdistalN2K.dat;

VARIABLE:    NAMES = u1-u12 y class1 class2;
             USEVARIABLES = u1-u12;
             CLASSES = c1(4) c2(4);
             CATEGORICAL = u1-u12;

ANALYSIS:    TYPE = MIXTURE;
             STARTS = 400 100;
             PROCESSORS = 12;

MODEL:       %OVERALL%
             c2 ON c1;

MODEL C1:    %C1#1%
             [u1$1-u6$1] (m1-m6);
             %C1#2%
             [u1$1-u6$1] (m21-m26);
             %C1#3%
             [u1$1-u6$1] (m31-m36);
             %C1#4%
             [u1$1-u6$1] (m41-m46);

MODEL C2:    %C2#1%
             [u7$1-u12$1] (m1-m6);
             %C2#2%
             [u7$1-u12$1] (m21-m26);
             %C2#3%
             [u7$1-u12$1] (m31-m36);
             %C2#4%
             [u7$1-u12$1] (m41-m46);
```

# Warnings for LTA Without Covariates and Distal Outcomes

THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES MAY NOT BE
TRUSTWORTHY FOR SOME PARAMETERS DUE TO A NON-POSITIVE DEFINITE
FIRST-ORDER DERIVATIVE PRODUCT MATRIX. THIS MAY BE DUE TO THE STARTING
VALUES BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION. THE
CONDITION NUMBER IS 0.406D-16. PROBLEM INVOLVING THE FOLLOWING PARAMETER:
Parameter 37, C2#1 ON C1#3

Parameter 37, C2#1 ON C1#3 DOES NOT AFFECT THE DATA FIT IN THE RANGE THE PARAMETER IS IN.
THE PARAMETER MIGHT HAVE CONVERGED TO A LARGE ABSOLUTE VALUE BEYOND WHICH THE MODEL
FIT DOES NOT IMPROVE. THE PROBLEM CAN OCCUR, FOR EXAMPLE, IF THE JOINT DISTRIBUTION
OF CATEGORICAL VARIABLES INVOLVED HAS EMPTY CELLS. TO RESOLVE THE PROBLEM, FIX THE
PARAMETER TO THE VALUE IT IS ESTIMATED AT.

ONE OR MORE MULTINOMIAL LOGIT 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 LATENT VARIABLES AND ANY INDEPENDENT
VARIABLES. THE FOLLOWING PARAMETERS WERE FIXED:
Parameter 33, C2#3 ON C1#1

- See also Web Talk 2, Error messages, Segments 21 and 22
  - C2#1 ON C1#3 is not fixed but flagged by the MLF information matrix check that suggests fixing it at the value estimated
  - C2#3 ON C1#1 is fixed based on the ML information matrix check

CATEGORICAL LATENT VARIABLES

| | Estimate | S.E. | Est./S.E. | Two-Tailed P-Value |
|---|---|---|---|---|
| C2#1 ON | | | | |
| C1#1 | 5.675 | 1.003 | 5.658 | 0.000 |
| C1#2 | 1.949 | 0.590 | 3.301 | 0.001 |
| C1#3 | **-14.738** | 1.256 | -11.732 | 0.000 |
| C2#2 ON | | | | |
| C1#1 | 5.564 | 1.086 | 5.122 | 0.000 |
| C1#2 | 4.859 | 0.625 | 7.780 | 0.000 |
| C1#3 | 4.730 | 1.255 | 3.768 | 0.000 |
| C2#3 ON | | | | |
| C1#1 | **-18.045** | 0.000 | **999.000** | 999.000 |
| C1#2 | 2.674 | 0.921 | 2.904 | 0.004 |
| C1#3 | 6.770 | 1.397 | 4.846 | 0.000 |

- C2#1 ON C1#3 is not fixed but flagged by the MLF information matrix check that suggests fixing it at the value estimated
- C2#3 ON C1#1 is fixed based on the ML information matrix check

# Joint Latent Class Probabilities

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES BASED ON THE ESTIMATED MODEL

Latent Class Pattern

| | | |
|---|---|---|
| 1 1 | 341.52647 | 0.17076 |
| 1 2 | 164.64637 | 0.08232 |
| **1 3** | **0.00000** | **0.00000** |
| 1 4 | 9.88470 | 0.00494 |
| 2 1 | 43.33560 | 0.02167 |
| 2 2 | 428.09734 | 0.21405 |
| 2 3 | 34.05369 | 0.01703 |
| 2 4 | 52.06235 | 0.02603 |
| **3 1** | **0.00000** | **0.00000** |
| 3 2 | 76.08995 | 0.03804 |
| 3 3 | 413.52924 | 0.20676 |
| 3 4 | 10.52070 | 0.00526 |
| 4 1 | 41.17109 | 0.02059 |
| 4 2 | 22.16256 | 0.01108 |
| 4 3 | 15.67266 | 0.00784 |
| 4 4 | 347.24728 | 0.17362 |

- Joint probability: $[C1, C2] = [C1] \times [C2|C1]$ where $[C2|C1]$ is the transition probability

# Latent Transition Probabilities for LTA Without Covariates and Distal Outcomes

LATENT TRANSITION PROBABILITIES BASED ON THE ESTIMATED MODEL

C1 Classes (Rows) by C2 Classes (Columns)

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | |
| 1 | 0.662 | 0.319 | **0.000** | 0.019 |
| 2 | 0.078 | 0.768 | 0.061 | 0.093 |
| 3 | **0.000** | 0.152 | 0.827 | 0.021 |
| 4 | 0.097 | 0.052 | 0.037 | 0.815 |

# First Step of Multistep Analysis Adding the Distal Outcome

```
TITLE:

DATA:        FILE = LTAEmptyCellsdistalN2K.dat;

VARIABLE:    NAMES = u1-u12 y class1 class2;
             USEVARIABLES = u1-u12;
             CLASSES = c1(4) c2(4);
             CATEGORICAL = u1-u12;
             AUXILIARY = y;

ANALYSIS:    TYPE = MIXTURE;
             STARTS = 400 100;
             PROCESSORS = 12;

MODEL:       %OVERALL%
             ! c2 ON c1;

MODEL C1:    %C1#1%
             [u1$1-u6$1] (m1-m6);
             %C1#2%
             [u1$1-u6$1] (m21-m26);
             %C1#3%
             [u1$1-u6$1] (m31-m36);
             %c1#4%
             [u1$1-u6$1] (m41-m46);

MODEL C2:    %C2#1%
             [u7$1-u12$1] (m1-m6);
             %C2#2%
             [u7$1-u12$1] (m21-m26);
             %C2#3%
             [u7$1-u12$1] (m31-m36);
             %C2#4%
             [u7$1-u12$1] (m41-m46);

SAVEDATA:    SAVE = CPROB BCHWEIGHTS;
             FILE = stageA.dat;

OUTPUT:      SVALUES;
```

# First Step of Multistep Analysis Adding the Distal Outcome: Reordering the Classes to Agree with LTA without Distal

```
TITLE:

DATA:          FILE = LTAEmptyCellsdistalN2K.dat;

VARIABLE:      NAMES = u1-u12 y class1 class2;
               USEVARIABLES = u1-u12;
               CLASSES = c1(4) c2(4);
               CATEGORICAL = u1-u12;
               AUXILIARY = y;

ANALYSIS:      TYPE = MIXTURE;
               OPTSEED = 629320;
MODEL:         %OVERALL%
               ! c2 ON c1;

MODEL C1:      %C1#1%
               [u1$1-u6$1] (m1-m6);
               %C1#2%
               [u1$1-u6$1] (m21-m26);
               %C1#3%
               [u1$1-u6$1] (m31-m36);
               %c1#4%
               [u1$1-u6$1] (m41-m46);

MODEL C2:      %C2#1%
               [u7$1-u12$1] (m1-m6);
               %C2#2%
               [u7$1-u12$1] (m21-m26);
               %C2#3%
               [u7$1-u12$1] (m31-m36);
               %C2#4%
               [u7$1-u12$1] (m41-m46);

SAVEDATA:      SAVE = CPROB BCHWEIGHTS;
               FILE = ReorderedstageA.dat;

OUTPUT:        SVALUES(1 3 4 2 | 1 3 4 2);
```

# Last Step of Multistep Analysis using BCH

```
TITLE:      Last step, BCH

DATA:       FILE = ReorderedstageA.dat;

VARIABLE:   NAMES = u1-u12 y w1-w16 cp1-cp16
            mlc1 mlc2 mlcjoint;
            USEVARIABLES = y w1-w16;
            CLASSES = c1(4) c2(4);
            TRAINING = w1-w16(BCH);

ANALYSIS:   TYPE = MIXTURE;
            STARTS = 0;

MODEL:      %OVERALL%
            c2 ON c1;

MODEL C2:   %C2#1%
            [y];
            %C2#2%
            [y];
            %C2#3%
            [y];
            %C2#4%
            [y];
```

# Latent Transition Probabilities from Last Step using BCH

C1 Classes (Rows) by C2 Classes (Columns)

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0.692 | 0.308 | **0.000** | **0.000** |
| 2 | 0.049 | 0.739 | 0.104 | 0.109 |
| 3 | **0.000** | 0.151 | 0.832 | 0.017 |
| 4 | 0.048 | 0.113 | 0.041 | 0.798 |

- Compare with slide 48: Cell 1,4 0.019 $\rightarrow$ 0.000
- Latent class probabilities:
    - Last step, BCH:
        - C1: 0.234, 0.268, 0.248, 0.250
        - C2: 0.187, 0.336, 0.244, 0.233
    - First step:
        - C1: 0.258, 0.279, 0.250, 0.213
        - C2: 0.213, 0.346, 0.232, 0.210

# Outline

- Regular LTA
- Multistep LTA with distal outcomes and covariates with missing data
    - 3-step, BCH, and 2-step using imputation
    - Inputs, results
- LTA with transition-specific distal outcome means, dot language
- Empty cells
    - **BCH failure**
- Multiple-group LTA
- RI-LTA
- LCA at two timepoints

- BCH weights:
  - Can be negative
  - Sum to 1 over classes for each individual
  - Should not sum to a negative value over individuals for any class:
    - Can prevent convergence - fatal error stoppage due to “total negative weight” - BCH fails
    - If convergence occurs, the solution is ok
    - Can happen with empty cells, or small class sizes and low entropy
    - Example: Distal outcome mean varying across all transition patterns with empty cells
- Solutions to BCH weight problems:
  - Estimated negative variances for continuous distal outcomes: Hold variances equal across classes (default)
  - Estimated negative frequencies for categorical distal outcomes: Hold probabilities equal across certain classes
  - Use another multistep method such as 2-step

# Distal Outcome Mean Varying Across All Transition Patterns: Last Step of BCH Using the Dot Language

```
TITLE:      Last step, BCH,
            dot language for distal means
            varying across transition patterns

DATA:       FILE= stageA.dat;

VARIABLE:   NAMES = u1-u12 y w1-w16 cp1-
            cp16 mlc1 mlc2 mlcjoint;
            USEVARIABLES = y w1-w16;
            CLASSES = c1(4) c2(4);
            TRAINING = w1-w16(bch);

ANALYSIS:   TYPE = MIXTURE;
            STARTS= 0;

MODEL:      %OVERALL%
            c2 ON c1;
            %c1#1.c2#1%
            [y];
            %c1#1.c2#2%
            [y];
            %c1#1.c2#3%
            [y];
            %c1#1.c2#4%
            [y];

            %c1#2.c2#1%
            [y];
            %c1#2.c2#2%
            [y];
            %c1#2.c2#3%
            [y];
            %c1#2.c2#4%
            [y];
            %c1#3.c2#1%
            [y];
            %c1#3.c2#2%
            [y];
            %c1#3.c2#3%
            [y];
            %c1#3.c2#4%
            [y];

            %c1#4.c2#1%
            [y];
            %c1#4.c2#2%
            [y];
            %c1#4.c2#3%
            [y];
            %c1#4.c2#4%
            [y];
```

# Distal Outcome Mean Varying Across All Transition Patterns: Error Message

THE ESTIMATED COVARIANCE MATRIX FOR THE Y VARIABLES IN CLASS 1 IS NOT POSITIVE DEFINITE. PROBLEM INVOLVING VARIABLE Y. COMPUTATION COULD NOT BE COMPLETED IN ITERATION 2. CHANGE YOUR MODEL AND/OR STARTING VALUES. THIS MAY BE DUE TO A ZERO ESTIMATED VARIANCE, THAT IS, NO WITHIN-CLASS VARIATION FOR THE VARIABLE.

THE WARNING IS DUE TO A NEGATIVE VARIANCE/RESIDUAL VARIANCE. VARIABLE Y HAS A NEGATIVE RESIDUAL VARIANCE. AVOID SPECIFYING CLASS-INVARIANT COVARIANCE WHEN CORRESPONDING VARIANCES ARE NOT SPECIFIED AS CLASS-INVARIANT.

*** FATAL ERROR THE TOTAL BCH WEIGHT FOR CLASS PATTERN 1 2 IS NEGATIVE: -8.662.
THIS IS LIKELY AN EMPTY CLASS. THE BCH MODEL ESTIMATION COULD NOT BE COMPLETED.

# BCH Failure

*** FATAL ERROR THE TOTAL BCH WEIGHT FOR CLASS PATTERN 1 2 IS NEGATIVE: -8.662. THIS IS LIKELY AN EMPTY CLASS. THE BCH MODEL ESTIMATION COULD NOT BE COMPLETED.

SAMPLE STATISTICS

| | | | | |
|---|---|---|---|---|
| Means | | | | |
| W1 | W2 | W3 | W4 | W5 |
| 0.169 | -0.004 | 0.077 | -0.007 | 0.009 |
| Means | | | | |
| W6 | W7 | W8 | W9 | W10 |
| 0.201 | 0.022 | 0.017 | 0.016 | 0.031 |
| Means | | | | |
| W11 | W12 | W13 | W14 | W15 |
| 0.195 | 0.025 | -0.001 | 0.005 | 0.037 |
| Means | | | | |
| W16 | | | | |
| 0.206 | | | | |

-8.662 = -0.004 * 2000

# Outline

- Regular LTA
- Multistep LTA with distal outcomes and covariates with missing data
  - 3-step, BCH, and 2-step using imputation
  - Inputs, results
- LTA with transition-specific distal outcome means, dot language
- Empty cells
  - BCH failure
- **Multiple-group LTA**
- RI-LTA
- LCA at two timepoints

# Multiple-Group LTA

- Example: Using LTADistalN2K.dat, the binary X1 variable is treated as the group variable
- X1 can be specified as a perfect indicator of a new latent class variable C in addition to C1, C2
    - KNOWNCLASS not allowed with BCH training variables
- This approach can be combined with 1-step, 2-step, 3-step, and BCH with or without imputation of missing data on x's.
- Model flexibility can be obtained using the dot language, specifying particular combinations of latent classes such as: %C#1.C1#1.C2#1%

  - Any parameter can be different, or held equal, across any combinations of latent classes

# Input for Multiple-Group Regular LTA

```
TITLE:        Regular 2-group LTA
DATA:         FILE = LTADistalN2K.dat;

VARIABLE:     NAMES = u1-u12 x1 y1 y2 x2 y3 x3 class1 class2;
              USEVARIABLES = u1-u12 x1;
              MISSING = ALL(999);
              CLASSES = c(2) c1(3) c2(3);
              CATEGORICAL = u1-u12 x1;

ANALYSIS:     TYPE = MIXTURE;
              STARTS = 40 10;
```

# Input for Multiple-Group Regular LTA, Continued

```
MODEL:        %OVERALL%
              c2 ON c1;
              c2 c1 ON c;

MODEL C:      %c#1%
              [x1$1@15]; ! P(x1=0)=1 for c class = 1
              %c#2%
              [x1$1@-15]; ! P(x1=1)=1 for c class = 2

MODEL C1:     %C1#1%
              [u1$1-u6$1] (1-6);

              %C1#2%
              [u1$1-u6$1] (11-16);

              %C1#3%
              [u1$1-u6$1] (21-26);

MODEL C2:     %C2#1%
              [u7$1-u12$1] (1-6);

              %C2#2%
              [u7$1-u12$1] (11-16);

              %C2#3%
              [u7$1-u12$1] (21-26);

OUTPUT:
```

# Input for Multiple-Group Multistep LTA
# First Step of BCH

```
TITLE:        First step of 2-group BCH LTA
              with x1 as group variable

DATA:         FILE = LTADistalN2K.dat;

VARIABLE:     NAMES = u1-u12 x1 y1 y2 x2
              y3 x3 class1 class2;
              USEVARIABLES = u1-u12 x1;
              MISSING = ALL(999);
              CLASSES = c(2) c1(3) c2(3);
              CATEGORICAL = u1-u12 x1;
              AUXILIARY = y1 y2 y3;

ANALYSIS:     TYPE = MIXTURE;
              STARTS = 40 10;

MODEL:        %OVERALL%
              ! c2 ON c1;
              ! c2 c1 ON c;

MODEL C:      %c#1%
              [x1$1@15]; ! P(x1=0)=1 for c class = 1
              %c#2%
              [x1$1@-15]; ! P(x1=1)=1 for c class = 2

MODEL C1:     %C1#1%
              [u1$1-u6$1] (1-6);
              %C1#2%
              [u1$1-u6$1] (11-16);
              %C1#3%
              [u1$1-u3$1] (21-23);
              [u4$1-u6$1] (24-26);

MODEL C2:     %C2#1%
              [u7$1-u12$1] (1-6);
              %C2#2%
              [u7$1-u12$1] (11-16);
              %C2#3%
              [u7$1-u9$1] (21-23);
              [u10$1-u12$1] (24-26);

SAVEDATA:     SAVE = BCHWEIGHTS;
              FILE = mgbch.dat;
```

# Input for Multiple-Group Multistep LTA Last Step of BCH

```
TITLE:      Last step of 2-group BCH LTA
            with x1 as group variable

DATA:       FILE = mgbch.dat;

VARIABLE:   NAMES = u1-u12 x1 y1-y3 w1-w18;
            USEVARIABLES = y1-y3;
            MISSING = *;
            CLASSES = c(2) c1(3) c2(3);
            CATEGORICAL = y1-y2;
            TRAINING = w1-w18(BCH);

ANALYSIS:   TYPE = MIXTURE;
            STARTS = 0;

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

            ! y thresholds and means vary across
            ! all 18 class combinations as the default
```

# BCH Failure: Error for Middle Category of Ordinal Distal Outcome Y2

WARNING: IN CLASS PATTERN 1 1 3 VARIABLE Y2 HAS A NEGATIVE FREQUENCY FOR CATEGORY 2.

WARNING: IN CLASS PATTERN 1 3 1 VARIABLE Y2 HAS A NEGATIVE FREQUENCY FOR CATEGORY 2.

WARNING: IN CLASS PATTERN 2 3 2 VARIABLE Y2 HAS A NEGATIVE FREQUENCY FOR CATEGORY 2.

THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO AN ERROR IN THE COMPUTATION. CHANGE YOUR MODEL AND/OR STARTING VALUES.

# Input for Multiple-Group Multistep LTA Last Step of 2-Step

```
TITLE:        Last step of 2-group 2-step LTA
              with x1 as group variable

DATA:         FILE = mgbch.dat;

VARIABLE:     NAMES = u1-u12 x1 y1-y3 w1-w18;
              USEVARIABLES = u1-u12 x1 y1-y3;
              MISSING = *;
              CLASSES = c(2) c1(3) c2(3);
              CATEGORICAL = u1-u12 x1 y1-y2;

ANALYSIS:     TYPE = MIXTURE;
              STARTS = 0;

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

MODEL C:      %C#1%
              [ x1$1@15 ];
              %C#2%
              [ x1$1@-15 ];

MODEL C1:     %C1#1%
              [ u1$1@1.37946 ] (1);
              [ u2$1@1.57905 ] (2);
              [ u3$1@1.38788 ] (3);
              [ u4$1@1.34298 ] (4);
              [ u5$1@1.25729 ] (5);
              [ u6$1@1.37010 ] (6);

              %C1#2%
              [ u1$1@1.76359 ] (7);
              [ u2$1@1.24351 ] (8);
              [ u3$1@1.31693 ] (9);
              [ u4$1@-1.66399 ] (10);
              [ u5$1@-1.39184 ] (11);
              [ u6$1@-1.56220 ] (12);

              %C1#3%
              [ u1$1@-1.34039 ] (13);
              [ u2$1@-1.30800 ] (14);
              [ u3$1@-1.21208 ] (15);
              [ u4$1@-1.27214 ] (16);
              [ u5$1@-1.30615 ] (17);
              [ u6$1@-1.29090 ] (18);
```

# Input for Multiple-Group Multistep LTA
# Last Step of 2-Step Continued

```
MODEL C2:     %C2#1%
              [ u7$1@1.37946 ] (1);
              [ u8$1@1.57905 ] (2);
              [ u9$1@1.38788 ] (3);
              [ u10$1@1.34298 ] (4);
              [ u11$1@1.25729 ] (5);
              [ u12$1@1.37010 ] (6);

              %C2#2%
              [ u7$1@1.76359 ] (7);
              [ u8$1@1.24351 ] (8);
              [ u9$1@1.31693 ] (9);
              [ u10$1@-1.66399 ] (10);
              [ u11$1@-1.39184 ] (11);
              [ u12$1@-1.56220 ] (12);

              %C2#3%
              [ u7$1@-1.34039 ] (13);
              [ u8$1@-1.30800 ] (14);
              [ u9$1@-1.21208 ] (15);
              [ u10$1@-1.27214 ] (16);
              [ u11$1@-1.30615 ] (17);
              [ u12$1@-1.29090 ] (18);
```

# Threshold Estimates for Distal Ordinal Y2: 2-Step vs BCH for Latent Class Pattern 1 1 3

| Thresholds | 2-step | BCH |
|---|---|---|
| Y2$1 | 1.413 | 1.179 |
| Y2$2 | 1.413 | 1.082 |

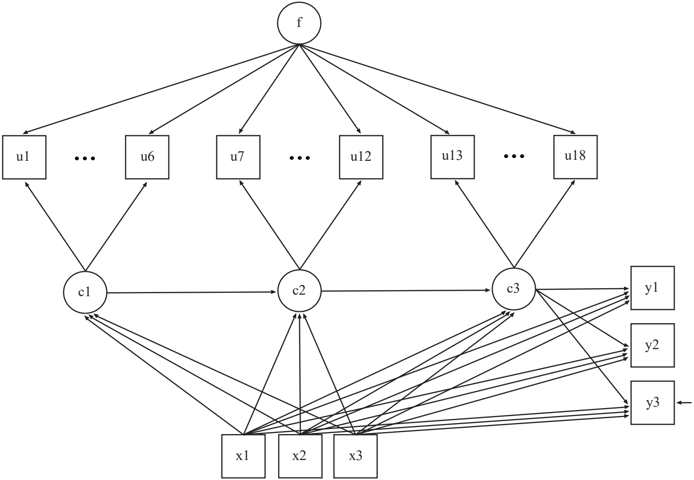
- 2-step thresholds collide - which is ok
  - Nobody in the middle Y2 category for latent class pattern 1 1 3
- BCH threshold 2 is lower than threshold 1 - which is not ok
  - BCH failure due to negative frequency

# Changing Ordinal Y2 to Nominal to Get BCH Results

```
TITLE:      Last step of 2-group BCH LTA
            with x1 as group variable

DATA:       FILE = mgbch.dat;

VARIABLE:   NAMES = u1-u12 x1 y1-y3 w1-w18;
            USEVARIABLES = y1-y3;
            MISSING = *;
            CLASSES = c(2) c1(3) c2(3);
            CATEGORICAL = y1;
            NOMINAL y2;
            TRAINING = w1-w18(BCH);

ANALYSIS:   TYPE = MIXTURE;
            STARTS = 0;

MODEL:      %OVERALL%
            c2 ON c1;
            c2 c1 ON c;
```

# Outline

- Regular LTA
- Multistep LTA with distal outcomes and covariates with missing data
  - 3-step, BCH, and 2-step using imputation
  - Inputs, results
- LTA with transition-specific distal outcome means, dot language
- Empty cells
  - BCH failure
- Multiple-group LTA
- **RI-LTA**
- LCA at two timepoints




- Muthén & Asparouhov (2022). Latent transition analysis with random intercepts (RI-LTA). Psychological Methods
- Web Talk 1

# Multistep Approaches for RI-LTA

- Simulations in Asparouhov & Muthén (Web Note 21):
  - Covariates
    - 2-step does better than BCH
  - Distal outcomes
    - BCH does better than 2-step when distals are non-normal within class
- The following analyses focus on multistep RI-LTA with distal outcomes (normal within class)
  - BCH
  - 2-step

# Input for RI-LTA: First Step of BCH

```
TITLE:      First step of RI-LTA using BCH

DATA:       FILE = RI-LTAdistalN2K.dat;

VARIABLE:   NAMES = u1-u18 x1 y1 y2 x2
            y3 x3 class1-class3;
            USEVARIABLES = u1-u18;
            CATEGORICAL = u1-u18;
            MISSING = ALL(999);
            CLASSES = c1(3) c2(3) c3(3);
            AUXILIARY = y1 y2 y3;

ANALYSIS:   TYPE = MIXTURE;
            PROCESSORS = 16;
            STARTS = 40 10;
            ALGORITHM = INTEGRATION;

MODEL:      %OVERALL%
            f BY u1-u6* (p1-p6)
            u7-u12 (p1-p6)
            u13-u18 (p1-p6);
            [f@0]; f@1;
            c2 ON c1; c3 ON c2;

MODEL C1:   %C1#1%
            [u1$1-u6$1] (1-6);
            %C1#2%
            [u1$1-u6$1] (7-12);
            %C1#3%
            [u1$1-u6$1] (13-18);

MODEL C2:   %C2#1%
            [u7$1-u12$1] (1-6);
            %C2#2%
            [u7$1-u12$1] (7-12);
            %C2#3%
            [u7$1-u12$1] (13-18);
```

# Input for RI-LTA: First Step of BCH, Continued

```
MODEL C3:      %C3#1%
               [u13$1-u18$1] (1-6);
               %C3#2%
               [u13$1-u18$1] (7-12);
               %C3#3%
               [u13$1-u18$1] (13-18);

SAVEDATA:      SAVE = BCHWEIGHTS;
               FILE = ri-ltabch.dat;
```

# Input for RI-LTA: Last Step of BCH

```
TITLE:      Last step of RI-LTA using BCH

DATA:       FILE = ri-ltabch.dat;

VARIABLE:   NAMES = u1-u18 y1-y3 w1-w27;
            USEVARIABLES = y1-y3;
            CATEGORICAL = y1 y2;
            MISSING = *;
            CLASSES = c1(3) c2(3) c3(3);
            TRAINING = w1-w27(BCH);
            ! Note: no F BY U so
            ! no ALGORITHM = INTEGRATION

ANALYSIS:   TYPE = MIXTURE;
            PROCESSORS = 16;
            STARTS = 0;

MODEL:      %OVERALL%
            c2 ON c1;
            c3 ON c2;

MODEL C3:   %C3#1%
            [y1$1 y2$1 y2$2 y3];
            %C3#2%
            [y1$1 y2$1 y2$2 y3];
            %C3#3%
            [y1$1 y2$1 y2$2 y3];
```

# Input for RI-LTA: First Step of 2-Step

```
TITLE:       First step of 2-step LTA

DATA:        FILE = RI-LTAdistalN2K.dat;

VARIABLE:    NAMES = u1-u18 x1 y1 y2 x2
             y3 x3 class1-class3;
             USEVARIABLES = u1-u18;
             CATEGORICAL = u1-u18;
             MISSING = ALL(999);
             CLASSES = c1(3) c2(3) c3(3);

ANALYSIS:    TYPE = MIXTURE;
             PROCESSORS = 16;
             STARTS = 40 10;
             ALGORITHM = INTEGRATION;

MODEL:       %OVERALL%
             f BY u1-u6* (p1-p6)
             u7-u12 (p1-p6)
             u13-u18 (p1-p6);
             [f@0]; f@1;
             c2 ON c1; c3 ON c2;

MODEL C1:    %C1#1%
             [u1$1-u6$1] (1-6);
             %C1#2%
             [u1$1-u6$1*-1.386] (7-12);
             %C1#3%
             [u1$1-u6$1] (13-18);

MODEL C2:    %C2#1%
             [u7$1-u12$1] (1-6);
             %C2#2%
             [u7$1-u12$1] (7-12);
             %C2#3%
             [u7$1-u12$1] (13-18);

MODEL C3:    %C3#1%
             [u13$1-u18$1] (1-6);
             %C3#2%
             [u13$1-u18$1] (7-12);
             %C3#3%
             [u13$1-u18$1] (13-18);

OUTPUT:      SVALUES;
```

# Input for RI-LTA: Last Step of 2-Step

```
TITLE:      Last step of 2-step RI-LTA

DATA:       FILE = RI-LTAdistalN2K.dat;

VARIABLE:   NAMES = u1-u18 x1 y1 y2 x2 y3 x3
            class1-class3;
            USEVARIABLES = u1-u18 y1 y2 y3;
            CATEGORICAL = u1-u18 y1 y2;
            MISSING = ALL(999);
            CLASSES = c1(3) c2(3) c3(3);

ANALYSIS:   TYPE = MIXTURE;
            PROCESSORS = 16;
            STARTS = 0;
            ALGORITHM = INTEGRATION;

MODEL:      %OVERALL%
            f BY u1;
            f BY u2;
            f BY u3;
            f BY u4;
            f BY u5;
            f BY u6;
            f BY u7;
            f BY u8;
            f BY u9;
            f BY u10;
            f BY u11;
            f BY u12;
            f BY u13;
            f BY u14;
            f BY u15;
            f BY u16;
            f BY u17;
            f BY u18;
```

# Input for RI-LTA: Last Step of 2-Step, Cont'd

```
c2#1 ON c1#1@1.10346;
c2#1 ON c1#2@0.47808;
c2#2 ON c1#1@0.72384;
c2#2 ON c1#2@1.28761;
c3#1 ON c2#1@0.58921;
c3#1 ON c2#2@-0.43383;
c3#2 ON c2#1@-0.30565;
c3#2 ON c2#2@0.46769;

[ c1#1@-0.35610 ];
[ c1#2@0.18106 ];
[ c2#1@-0.72994 ];
[ c2#2@-0.36206 ];
[ c3#1@-0.28953 ];
[ c3#2@0.22931 ];
```

```
f BY u1@0.85788 (p1);
f BY u2@1.06903 (p2);
f BY u3@0.97320 (p3);
f BY u4@0.95049 (p4);
f BY u5@0.98050 (p5);
f BY u6@0.91785 (p6);
f BY u7@0.85788 (p1);
f BY u8@1.06903 (p2);
f BY u9@0.97320 (p3);
f BY u10@0.95049 (p4);
f BY u11@0.98050 (p5);
f BY u12@0.91785 (p6);
f BY u13@0.85788 (p1);
f BY u14@1.06903 (p2);
f BY u15@0.97320 (p3);
f BY u16@0.95049 (p4);
f BY u17@0.98050 (p5);
f BY u18@0.91785 (p6);

[ f@0 ];
f@1;
```

# Input for RI-LTA: Last Step of 2-Step, Cont'd

```
MODEL C1:   %C1#1%

            [ u1$1@-1.59889 ] (7);
            [ u2$1@-1.46678 ] (8);
            [ u3$1@-1.92957 ] (9);
            [ u4$1@-1.35463 ] (10);
            [ u5$1@-1.51005 ] (11);
            [ u6$1@-1.50796 ] (12);

            %C1#2%

            [ u1$1@1.28855 ] (13);
            [ u2$1@1.40279 ] (14);
            [ u3$1@1.26994 ] (15);
            [ u4$1@1.25817 ] (16);
            [ u5$1@1.50531 ] (17);
            [ u6$1@1.38278 ] (18);

            %C1#3%

            [ u1$1@1.12320 ] (19);
            [ u2$1@1.19771 ] (20);
            [ u3$1@1.41476 ] (21);
            [ u4$1@-1.16953 ] (22);
            [ u5$1@-1.34632 ] (23);
            [ u6$1@-1.27857 ] (24);

MODEL C2:   %C2#1%

            [ u7$1@-1.59889 ] (7);
            [ u8$1@-1.46678 ] (8);
            [ u9$1@-1.92957 ] (9);
            [ u10$1@-1.35463 ] (10);
            [ u11$1@-1.51005 ] (11);
            [ u12$1@-1.50796 ] (12);
```

```
%C2#2%

[ u7$1@1.28855 ] (13);
[ u8$1@1.40279 ] (14);
[ u9$1@1.26994 ] (15);
[ u10$1@1.25817 ] (16);
[ u11$1@1.50531 ] (17);
[ u12$1@1.38278 ] (18);

%C2#3%

[ u7$1@1.12320 ] (19);
[ u8$1@1.19771 ] (20);
[ u9$1@1.41476 ] (21);
[ u10$1@-1.16953 ] (22);
[ u11$1@-1.34632 ] (23);
[ u12$1@-1.27857 ] (24);

MODEL C3: %C3#1%

[ u13$1@-1.59889 ] (7);
[ u14$1@-1.46678 ] (8);
[ u15$1@-1.92957 ] (9);
[ u16$1@-1.35463 ] (10);
[ u17$1@-1.51005 ] (11);
[ u18$1@-1.50796 ] (12);
[y1$1 y2$1 y2$2 y3];

%C3#2%

[ u13$1@1.28855 ] (13);
[ u14$1@1.40279 ] (14);
[ u15$1@1.26994 ] (15);
[ u16$1@1.25817 ] (16);
[ u17$1@1.50531 ] (17);
[ u18$1@1.38278 ] (18);
[y1$1 y2$1 y2$2 y3];
```

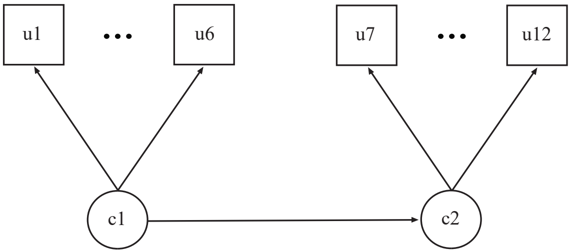
# Input for RI-LTA: Last Step of 2-Step, Cont'd

```
%C3#3%

[ u13$1@1.12320 ] (19);
[ u14$1@1.19771 ] (20);
[ u15$1@1.41476 ] (21);
[ u16$1@-1.16953 ] (22);
[ u17$1@-1.34632 ] (23);
[ u18$1@-1.27857 ] (24);
[y1$1 y2$1 y2$2 y3];
```

# Outline

- Regular LTA
- Multistep LTA with distal outcomes and covariates with missing data
  - 3-step, BCH, and 2-step using imputation
  - Inputs, results
- LTA with transition-specific distal outcome means, dot language
- Empty cells
  - BCH failure
- Multiple-group LTA
- RI-LTA
- **LCA at two timepoints**

# LCA at Two Timepoints




- The two timepoints can have different latent class constructs, different types of latent class indicators, and different number of latent classes
- Interest in predicting the C2 classes from the C1 classes
- Desire to not have timepoint 2 indicators influence the timepoint 1 class formation and vice versa
  - BCH multistep approach inspired by LTA:
    - First step: No measurement invariance, no C2 on C1
    - Last step: No observed variables, C2 on C1

# First Step of BCH for LCA at Two Timepoints

```
TITLE:      First BCH step of 2 LCAs

DATA:       FILE = LTAdistalN2K.dat;

VARIABLE:   NAMES = u1-u12 x1 y1 y2 x2
            y3 x3 class1 class2;
            USEVARIABLES = u1-u12;
            MISSING = ALL(999);
            CLASSES = c1(3) c2(3);
            CATEGORICAL = u1-u12;

ANALYSIS:   TYPE = MIXTURE;
            STARTS = 80 20;

MODEL:      %OVERALL%

MODEL C1:   %C1#1%
            [u1$1-u6$1];
            %C1#2%
            [u1$1-u6$1];
            %C1#3%
            [u1$1-u6$1];

MODEL C2:   %C2#1%
            [u7$1-u12$1];
            %C2#2%
            [u7$1-u12$1];
            %C2#3%
            [u7$1-u12$1];

SAVEDATA:   SAVE = BCHWEIGHTS;
            FILE = bch2LCA.dat;
```

# Last Step of BCH for LCA at Two Timepoints

```
TITLE:       Last BCH step of 2 LCAs
DATA:        FILE = bch2LCA.dat;

VARIABLE:    NAMES = u1-u12 w1-w9;
             USEVARIABLES = w1-w9;
             MISSING = *;
             CLASSES = c1(3) c2(3);
             TRAINING = w1-w9(BCH);

ANALYSIS:    TYPE = MIXTURE;
             STARTS = 0;

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

# Appendix: Monte Carlo Generation of LTA

```
TITLE:            Distal Y and missing on X in LTA
                  x1: binary, missing
                  x2: continuous, missing
                  x3: continuous, not missing
                  y1: binary, not missing
                  y2: ordinal, not missing
                  y3: continuous, not missing

MONTECARLO:       NAMES = u1-u12 x1 x2 x3 y1 y2 y3;
                  NOBSERVATIONS = 2000;
                  NREPS = 1;
                  SAVE = LTAdistalN2K.dat;
                  CLASSES = c1(3) c2(3);
                  GENCLASSES = c1(3) c2(3);
                  GENERATE = u1-u12(1) x1(1) y1(1) y2(2);
                  CATEGORICAL = u1-u12 x1 y1 y2;
                  MISSING = x1 x2;
                  ! seed = 347;

MODEL MISSING:    %OVERALL%
                  [x1*-1.5 x2*-1.5]; x1 x2 ON x3*0.4;
```

# Appendix: Monte Carlo Generation of LTA, Continued

```
ANALYSIS:              TYPE = MIXTURE;
                       ALGORITHM = INTEGRATION;
                       INTEGRATION = MONTECARLO;

MODEL POPULATION:      %OVERALL%
                       [x1$1*0]; x1 ON x3*1; x3*1;
                       [x2*0]; x2 ON x3*1; x2*1;
                       c1#1 c2#1 ON x1*0.5 x2*-0.3 x3*0.2;
                       c1#2 c2#2 ON x1*-.5 x2*.3 x3*-.2;
                       c2#1 ON c1#1*0.5; c2#2 ON c1#1*0;
                       c2#1 ON c1#2*0; c2#2 ON c1#2*0.5;

                       y1-y3 ON x1-x3*0.5;
                       y3*1;
```

# Appendix: Monte Carlo Generation of LTA, Continued

```
MODEL
POPULATION-
C1:
            %C1#1%
            [u1$1-u6$1*1.386];
            ! [y1$1*0.5 y2$1*0.2 y2$2*0.4 y3*1];

            %C1#2%
            [u1$1-u6$1*-1.386];
            ! [y1$1*-0.5 y2$1*-0.2 y2$2*0 y3*-1];

            %C1#3%
            [u1$1-u3$1*1.386];
            [u4$1-u6$1*-1.386];
            ! [y1$1*0.5 y2$1*-0.2 y2$2*0.2 y3*0];
```

```
MODEL
POPULATION-
C2:
            %C2#1%
            [u7$1-u12$1*1.386];
            [y1$1*0.5 y2$1*0.2 y2$2*0.4 y3*1];

            %C2#2%
            [u7$1-u12$1*-1.386];
            [y1$1*-0.5 y2$1*-0.2 y2$2*0 y3*-1];

            %C2#3%
            [u7$1-u9$1*1.386];
            [u10$1-u12$1*-1.386];
            [y1$1*0.5 y2$1*-0.2 y2$2*0.2 y3*0];
```

Repeat for MODEL C1, MODEL C2

# 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.
- Bakk, Tekle, & Vermunt (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. Sociological Methodology, 43, 272-311.
- Muthén & Asparouhov (2022). Latent transition analysis with random intercepts (RI-LTA). Psychological Methods, 27(1), 1–16