Analysis of causal effects with M*plus*

Axel Mayer[1,2]

[1]Ghent University, [2]Friedrich Schiller University Jena

July 21st, 2014

Introduction

The Big Picture

- We are interested in causal effects of a treatment (cause) X on an outcome Y.
- We are not only interested in the average causal effect, but also in conditional causal effects (e.g., effects on the treated, effects for patients with a high neediness, effects for females,...).
- We compute causal effects based on an unbiased regression $E(Y \mid X, K, Z)$, where K are categorical covariates and Z are continuous covariates.
- The covariates must be selected in such a way, that unbiasedness of $E(Y \mid X, K, Z)$ holds. While unbiased itself is not testable, some of the (stronger) causality conditions implying unbiasedness can be tested empirically.
- Based on the unbiased regression $E(Y|X, K, Z)$, we compute average and conditional causal effects using a multigroup SEM approach.

We present a multigroup structural equation modeling approach that allows for the identification of average and conditional effects of a discrete treatment variable X on a continuous outcome variable Y, taking into account categorical covariates K and continuous covariates Z.

The treatment variable X can take on values $0, 1, \ldots, T$ and the regression $E(Y \mid X, K, Z)$ can always be parameterized as follows:

$$E(Y \mid X, K, Z) = g_0(K, Z) + g_1(K, Z) \cdot I_{X=1} + \ldots + g_T(K, Z) \cdot I_{X=T},$$

where $g_0(K, Z), \ldots, g_T(K, Z)$ are any functions of K, Z and $I_{X=t}$ is an indicator variable for $X = t$. We call the $X = 0$ the 'control group' and all other treatment groups $X = 1, \ldots, T$ are compared to it.

Average and conditional effects:

The values of the $g_t(K,Z)$-functions are the conditional treatment effects of treatment $X=t$ vs. treatment $X=0$ given values of K and Z.

We can consider expectations and conditional expectations of the $g_t(K,Z)$-functions, i.e., average and conditional effects:

$E[g_t(K,Z)]$	Average effect (AVE)
$g_t(k,z)$	Conditional effect given a value z of Z and a value k of K
$E[g_t(K,Z) \mid X=t^*]$	Effect given $X=t^*$ (Effect on the treated)
$E[g_t(K,Z) \mid X=0]$	Effect on the non-treated
$E[g_t(K,Z) \mid K=k]$	Effect given a value $K=k$
$E[g_t(K,Z) \mid Z=z]$	Effect given a value $Z=z$
$E[g_t(K,Z) \mid V=v]$	Effect given a value $V=v$, where $V=f(X,K,Z)$

Multigroup Structural Equation Models

We will use multigroup structural equation models to estimate these average and conditional effects. The approach uses as many groups as there are combinations of treatment groups and levels of categorical covariates. Then, the regressions of the dependent variable on the continuous covariates are estimated in each cell and adjusted means as well as average and conditional causal effects are calculated.

The multigroup structural equation modeling approach allows for

1. latent covariates and latent outcome variables,
2. (higher order) interactions between the treatment variable and categorical and continuous covariates,
3. stochastic and fixed categorical and continuous covariates, and
4. heterogeneous residual variances across groups.

The multigroup structural equation model consists of a group-specific measurement model and a group-specific structural model:

$$\begin{pmatrix}\mathbf{Y}\\ \mathbf{Z}\end{pmatrix} = \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g \boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad \text{Measurement model for group } g$$

$$\boldsymbol{\eta} = \boldsymbol{\alpha}_g + \mathbf{B}_g \boldsymbol{\eta} + \boldsymbol{\zeta}_w \qquad \text{Structural model for group } g$$

The multigroup SEM implies a certain structure for means and (co-)variances of observed variables and the parameters of the model are estimated using ML (or other estimators).

In order to compute average effects, we also need a model for the group sizes (KNOWNCLASS option in M*plus*):

$$P(X=m) = \frac{\exp(\lambda_m)}{1 + \sum_{i=0}^{n-1} \exp(\lambda_i)} \qquad 0 \leq m \leq n-1$$

$$P(X=n) = 1 - \sum_{m=0}^{n-1} P(X=m)$$

Example 1

Generalized ANCOVA with two treatment conditions, linear g-functions, and a single continuous covariate Z:

$$\begin{aligned} E(Y \mid X, Z) &= g_0(Z) + g_1(Z) \cdot X \\ &= (\gamma_{00} + \gamma_{01} \cdot Z) + (\gamma_{10} + \gamma_{11} \cdot Z) \cdot X. \end{aligned}$$

Some average and conditional effects:

$E[g_1(Z)] = \gamma_{10} + \gamma_{11} \cdot E(Z)$	Average effect (AVE)
$g_1(z) = \gamma_{10} + \gamma_{11} \cdot z$	Conditional effect give a value z of Z
$E[g_1(Z) \mid X = 1] = \gamma_{10} + \gamma_{11} \cdot E(Z \mid X = 1)$	Effect on the treated
$E[g_1(Z) \mid X = 0] = \gamma_{10} + \gamma_{11} \cdot E(Z \mid X = 0)$	Effect on the non-treated

Example 1

Multigroup SEM:

$$\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} \beta_{00} \\ \mu_{01} \end{pmatrix} + \begin{pmatrix} \beta_{01} \\ 0 \end{pmatrix} Z + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} \qquad \text{Structural model } X = 0$$

$$\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \mu_{11} \end{pmatrix} + \begin{pmatrix} \beta_{11} \\ 0 \end{pmatrix} Z + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} \qquad \text{Structural model } X = 1$$

$$P(X=0) = \frac{\exp(\lambda_0)}{1+\exp(\lambda_0)} \qquad \text{Group size } X = 0$$

$$P(X=1) = 1 - P(X=0) \qquad \text{Group size } X = 1$$

(Measurement model is omitted, because there are no latent variables.)

Average and conditional effects as functions of model parameters:

$$\begin{aligned} g1(z) &= (\beta_{10} - \beta_{00}) + (\beta_{11} - \beta_{01}) \cdot z = \gamma_{10} + \gamma_{11} \cdot z \\ E[g1(Z) \mid X = 1] &= (\beta_{10} - \beta_{00}) + (\beta_{11} - \beta_{01}) \cdot \mu_{11} \\ E(Z) &= \mu_{01} \cdot P(X=0) + \mu_{11} \cdot P(X=1) \\ E[g1(Z)] &= (\beta_{10} - \beta_{00}) + (\beta_{11} - \beta_{01}) \cdot E(Z) \end{aligned}$$

M*plus* syntax for example 1:

```
TITLE:
   Example 01 with continuous covariate

DATA:
  file is data01.dat;
  type is individual;

VARIABLE:
  names are y x z;
  classes = cell(2);
  knownclass = cell(x=0 x=1);

ANALYSIS:
  type = mixture;
```

M*plus* syntax for example 1 (continued):

```
MODEL:

%overall%
  y on z; ! regression of y on z
  [cell#1] (lc1); ! coefficient for group size 0

%cell#1%
  y ON z* (b10); ! regression y on z in group x=0
  [y*] (b00); ! name for intercept of this regression
  [z*] (m0); ! conditional mean of z  in group x=0

%cell#2%
  y ON z* (b11);  ! regression y on z in group x=1
  [y*] (b01); ! name for intercept of this regression
  [z*] (m1); ! conditional mean of z  in group x=1
```

Example 1

M*plus* syntax for example 1 (continued):

```
MODEL CONSTRAINT:
 new(ga10 ga11); ! coefficients of effect function
 new(px0 px1 mz); ! P(X=0), P(X=1) and unconditional mean of z
 new(ae aet aeut ep1sd em1sd); ! various effects (see below)

 ga10 = b01-b00;
 ga11 = b11-b10;

 px0 = exp(lc1)/(1+exp(lc1));
 px1 = 1-px0;
 mz = m0*px0 + m1*px1;

 ae = ga10 + ga11*mz; ! average effect
 aet = ga10 + ga11*m1; ! average effect on the treated
 aeut = ga10 + ga11*m0; ! average effect on the untreated
 ep1sd = ga10 + ga11*(mz + 2); ! effect given mean z + 1 SD
 em1sd = ga10 + ga11*(mz - 2); ! effect given mean z - 1 SD
```

Example 2

Table 1.4. Conditional Expectations Given Treatment and Neediness

	Neediness						
Treatment	Low (Z=0)		Medium (Z=1)		High (Z=2)		
X=0	120	(20/120)	110	(17/120)	60	(3/120)	(40/120)
X=1	100	(7/120)	100	(26/120)	100	(7/120)	(40/120)
X=2	80	(3/120)	90	(17/120)	140	(20/120)	(40/120)
		(30/120)		(60/120)		(30/120)	

Note. Probabilities $P(X{=}x, Z{=}z)$, $P(Z{=}z)$, and $P(X{=}x)$ in parentheses.

Generalized ANCOVA with three treatment conditions and a single categorical covariate K with three values:

$$E(Y \mid X, K) = g_0(K) + g_1(K) \cdot I_{X=1} + g_2(K) \cdot I_{X=2}$$

$$\begin{aligned}
g_0(K) &= \gamma_{000} + \gamma_{010} \cdot I_{K=1} + \gamma_{020} \cdot I_{K=2} \\
g_1(K) &= \gamma_{100} + \gamma_{110} \cdot I_{K=1} + \gamma_{120} \cdot I_{K=2} \\
g_2(K) &= \gamma_{200} + \gamma_{210} \cdot I_{K=1} + \gamma_{220} \cdot I_{K=2}.
\end{aligned}$$

Some average and conditional effects:

$E[g_1(K)] = \gamma_{100} + \gamma_{110} \cdot P(K=1) + \gamma_{120} \cdot P(K=2)$	Average effect 1 vs. 0
$E[g_2(K)] = \gamma_{200} + \gamma_{210} \cdot P(K=1) + \gamma_{220} \cdot P(K=2)$	Average effect 2 vs. 0
$g_1(0) = \gamma_{100}$	Conditional effect 1 vs. 0 given $K=0$
$g_2(2) = \gamma_{200} + \gamma_{220}$	Conditional effect 2 vs. 0 given $K=2$
$E[g_1(K) \mid X=1] = \gamma_{100} + \gamma_{110} \cdot P(K=1 \mid X=1) +$ $\gamma_{120} \cdot P(K=2 \mid X=1)$	Effect 1 vs. 0 on the treated

Multigroup SEM:

$$\left(Y\right) = \left(\alpha_{xk0}\right) + \left(\zeta_1\right)$$ Structural model for each cell $C = c$

$$P(C = m) = \frac{\exp(\lambda_m)}{1 + \sum_{i=1}^{8} \exp(\lambda_i)} \qquad m = 1, 2, \ldots, 8$$ Group sizes

$$P(C = 9) = 1 - \sum_{m=1}^{8} P(C = m)$$ Group size reference group

Example 2

In an intermediate step, we compute the regression coefficients of $E^{X=x}(Y \mid K) = \beta_{x00} + \beta_{x10} \cdot I_{K=1} + \beta_{x20} \cdot I_{K=2}$ and then the coefficients for the effect functions (see below).

Computations for some average and conditional effects as functions of model parameters (only shown for $g_1(K)$, but similar for $g_2(K)$ – see M*plus* syntax for details):

$$
\begin{aligned}
\gamma_{100} &= \beta_{100} - \beta_{000} = \alpha_{100} - \alpha_{000} \\
\gamma_{110} &= \beta_{110} - \beta_{010} = (\alpha_{110} - \alpha_{100}) - (\alpha_{010} - \alpha_{000}) \\
\gamma_{120} &= \beta_{120} - \beta_{020} = (\alpha_{120} - \alpha_{100}) - (\alpha_{020} - \alpha_{000}) \\
P(K=1) &= P(X=0, K=1) + P(X=1, K=1) + P(X=2, K=1) \\
P(K=2) &= P(X=0, K=2) + P(X=1, K=2) + P(X=2, K=2) \\
E[g_1(K)] &= \gamma_{100} + \gamma_{110} \cdot P(K=1) + \gamma_{120} \cdot P(K=2) \\
g_1(0) &= \gamma_{100}
\end{aligned}
$$

Example 2

M*plus* syntax for example 2:

```
TITLE: Example 02: nonortho treat and neediness
DATA: FILE IS data02.dat;
DEFINE:
! compute knowclass variable C
IF K == 0 AND X == 0 THEN C = 1 ;
IF K == 1 AND X == 0 THEN C = 2 ;
IF K == 2 AND X == 0 THEN C = 3 ;
IF K == 0 AND X == 1 THEN C = 4 ;
IF K == 1 AND X == 1 THEN C = 5 ;
IF K == 2 AND X == 1 THEN C = 6 ;
IF K == 0 AND X == 2 THEN C = 7 ;
IF K == 1 AND X == 2 THEN C = 8 ;
IF K == 2 AND X == 2 THEN C = 9 ;

VARIABLE: NAMES = Y X K ; ! variable names
  USEVARIABLES = Y C;
      CLASSES = cell( 9 ); ! 9 cells (3x3 combinations of X and K)
  KNOWNCLASS = cell( C=1 C=2 C=3 C=4 C=5 C=6 C=7 C=8 C=9 );
ANALYSIS: TYPE = MIXTURE;
```

M*plus* syntax for example 2 (continued):

```
MODEL:

%OVERALL%
Y; ! We just estimate mean and variance of Y in each cell

! coefficients for group sizes 1-8 (logistic metric)
! group 9 is the reference group
[cell#1] (LC1);
[cell#2] (LC2);
[cell#3] (LC3);
[cell#4] (LC4);
[cell#5] (LC5);
[cell#6] (LC6);
[cell#7] (LC7);
[cell#8] (LC8);
```

Example 2

M*plus* syntax for example 2 (continued):

```
! mean of Y in each group
%cell#1%
[Y] (a0_0_0); ! intercept/mean of Y
%cell#2%
[Y] (a0_1_0);
%cell#3%
[Y] (a0_2_0);
%cell#4%
[Y] (a1_0_0);
%cell#5%
[Y] (a1_1_0);
%cell#6%
[Y] (a1_2_0);
%cell#7%
[Y] (a2_0_0);
%cell#8%
[Y] (a2_1_0);
%cell#9%
[Y] (a2_2_0);
```

Example 2

M*plus* syntax for example 2 (continued):

```
MODEL CONSTRAINT:

NEW(PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9 ! P(C=c)
PK1 PK2 ! P(K=k)
b0_0_0 b0_1_0 b0_2_0
b1_0_0 b1_1_0 b1_2_0
b2_0_0 b2_1_0 b2_2_0
g1_0_0 g1_1_0 g1_2_0 ! coefficients of effect function g1
g2_0_0 g2_1_0 g2_2_0 ! coefficients of effect function g2
Eg1 Eg2 ! average effect 1 vs 0 and average effect 2 vs. 0
Eg1gK0 Eg1gK1 Eg1gK2 ! effect 1 vs 0 given values k of K
Eg2gK0 Eg2gK1 Eg2gK2 ! effect 2 vs 0 given values k of K
);
```

M*plus* syntax for example 2 (continued):

```
! P(C=c)
PC1 = EXP(LC1) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
+EXP(LC7) +EXP(LC8) +1);
PC2 = EXP(LC2) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
+EXP(LC7) +EXP(LC8) +1);
PC3 = EXP(LC3) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
+EXP(LC7) +EXP(LC8) +1);
PC4 = EXP(LC4) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
+EXP(LC7) +EXP(LC8) +1);
PC5 = EXP(LC5) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
+EXP(LC7) +EXP(LC8) +1);
PC6 = EXP(LC6) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
+EXP(LC7) +EXP(LC8) +1);
PC7 = EXP(LC7) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
+EXP(LC7) +EXP(LC8) +1);
PC8 = EXP(LC8) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
+EXP(LC7) +EXP(LC8) +1);
PC9 = 1 - (PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8);
```

Example 2

M*plus* syntax for example 2 (continued):

```
! coefficients of (X=x)-conditional regressions of Y on K
! e.g. E^{X=0}(Y|K) = b0_0_0 + b0_1_0 * I_{K=1} + b0_2_0 * I_{K=2}
b0_0_0 = a0_0_0;
b0_1_0 = a0_1_0 - a0_0_0;
b0_2_0 = a0_2_0 - a0_0_0;
b1_0_0 = a1_0_0;
b1_1_0 = a1_1_0 - a1_0_0;
b1_2_0 = a1_2_0 - a1_0_0;
b2_0_0 = a2_0_0;
b2_1_0 = a2_1_0 - a2_0_0;
b2_2_0 = a2_2_0 - a2_0_0;

! coefficients of effect functions g1 and g2
g1_0_0 = b1_0_0 - b0_0_0;
g1_1_0 = b1_1_0 - b0_1_0;
g1_2_0 = b1_2_0 - b0_2_0;
g2_0_0 = b2_0_0 - b0_0_0;
g2_1_0 = b2_1_0 - b0_1_0;
g2_2_0 = b2_2_0 - b0_2_0;
```

Example 2

M*plus* syntax for example 2 (continued):

```
! P(K=k)
PK1 = PC2 + PC5 + PC8;
PK2 = PC3 + PC6 + PC9;

! effects
Eg1 = g1_0_0 + g1_1_0*PK1 + g1_2_0*PK2; ! average effect 1 vs. 0
Eg2 = g2_0_0 + g2_1_0*PK1 + g2_2_0*PK2; ! average effect 2 vs. 0

Eg1gK0 = g1_0_0; ! effect 1 vs. 0 given K=0
Eg1gK1 = g1_0_0 + g1_1_0; ! effect 1 vs. 0 given K=1
Eg1gK2 = g1_0_0 + g1_2_0; ! effect 1 vs. 0 given K=2

Eg2gK0 = g2_0_0; ! effect 2 vs. 0 given K=0
Eg2gK1 = g2_0_0 + g2_1_0; ! effect 2 vs. 0 given K=1
Eg2gK2 = g2_0_0 + g2_2_0; ! effect 2 vs. 0 given K=2
```

Example 3

Example 3:

- η_2: latent outcome variable with three indicators Y_{12}, Y_{22}, Y_{32}
- X: treatment variable (control, treat1, treat2)
- η_1: latent continuous covariate (pre-test) with three indicators Y_{11}, Y_{21}, Y_{31}
- K: gender (male, female)

$$E(\eta_2 \mid X, K, \eta_1) = g_0(K, \eta_1) + g_1(K, \eta_1) \cdot I_{X=1} + g_2(K, \eta_1) \cdot I_{X=2}$$

$$g_0(K, \eta_1) = \gamma_{000} + \gamma_{010} \cdot K + \gamma_{001} \cdot \eta_1 + \gamma_{011} \cdot K \cdot \eta_1$$
$$g_1(K, \eta_1) = \gamma_{100} + \gamma_{110} \cdot K + \gamma_{101} \cdot \eta_1 + \gamma_{111} \cdot K \cdot \eta_1$$
$$g_2(K, \eta_1) = \gamma_{200} + \gamma_{210} \cdot K + \gamma_{201} \cdot \eta_1 + \gamma_{211} \cdot K \cdot \eta_1$$

Example 3

Effect functions (repeated):

$$g_1(K,\eta_1) = \gamma_{100} + \gamma_{110} \cdot K + \gamma_{101} \cdot \eta_1 + \gamma_{111} \cdot K \cdot \eta_1$$
$$g_2(K,\eta_1) = \gamma_{200} + \gamma_{210} \cdot K + \gamma_{201} \cdot \eta_1 + \gamma_{211} \cdot K \cdot \eta_1$$

Some average and conditional effects:

Average effect of $X = 1$ vs. $X = 0$:

$$E[g_1(K,\eta_1)] = \gamma_{100} + \gamma_{110} \cdot E(K) + \gamma_{101} \cdot E(\eta_1) + \gamma_{111} \cdot E(K \cdot \eta_1)$$

Average effect of $X = 2$ vs. $X = 0$:

$$E[g_2(K,\eta_1)] = \gamma_{200} + \gamma_{210} \cdot E(K) + \gamma_{201} \cdot E(\eta_1) + \gamma_{211} \cdot E(K \cdot \eta_1)$$

Conditional effect of $X = 1$ vs. $X = 0$ given $K = 0$:

$$E[g_1(K,\eta_1) \mid K = 0] = \gamma_{100} + \gamma_{101} \cdot E(\eta_1 \mid K = 0)$$

Conditional effect of $X = 1$ vs. $X = 0$ given $X = 2$:

$$E[g_1(K,\eta_1) \mid X = 2] = \gamma_{100} + \gamma_{110} \cdot E(K \mid X = 2) + \gamma_{101} \cdot E(\eta_1 \mid X = 2) + \gamma_{111} \cdot E(K \cdot \eta_1 \mid X = 2)$$

Example 3

Multigroup SEM: Group-invariant τ-equivalent measurement model

$$\begin{pmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{12} \\ Y_{22} \\ Y_{32} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \end{pmatrix}$$

Multigroup SEM: Group-specific structural model in each of the six groups (2 x 3)

$$\begin{pmatrix} \eta_2 \\ \eta_1 \end{pmatrix} = \begin{pmatrix} \alpha_{xk0} \\ \mu_{xk1} \end{pmatrix} + \begin{pmatrix} \alpha_{xk1} \\ 0 \end{pmatrix} \eta_1 + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}$$

Multigroup SEM: Group weights for each of the six groups (2 x 3)

$$P(C=m) = \frac{\exp(\lambda_m)}{1 + \sum_{i=1}^{5} \exp(\lambda_i)} \qquad m = 1, 2, \ldots, 5$$

Group sizes

$$P(C=6) = 1 - \sum_{m=1}^{5} P(C=m)$$

Group size reference group

Based on the parameters of the model (previous slide), we can estimate average and conditional effects. I am not showing all details here but see M*plus* syntax.

Some examples:

$$\gamma_{100} = \alpha_{100} - \alpha_{000}$$
$$\gamma_{101} = \alpha_{101} - \alpha_{001}$$
$$\gamma_{110} = (\alpha_{110} - \alpha_{100}) - (\alpha_{010} - \alpha_{000})$$
$$\gamma_{111} = (\alpha_{111} - \alpha_{101}) - (\alpha_{011} - \alpha_{000})$$

$$E(\eta_1) = \mu_{001} \cdot P(X=0, K=0) + \mu_{011} \cdot P(X=0, K=1) + \ldots + \mu_{211} \cdot P(X=2, K=1)$$
$$P(K=1) = P(X=0, K=1) + P(X=1, K=1) + P(X=2, K=1)$$
$$P(X=0) = P(X=0, K=0) + P(X=0, K=1) + P(X=0, K=2)$$

M*plus* syntax for example 3:

```
TITLE: Example 03 with latent variables and categorical covariate
DATA: FILE IS data03.dat;
DEFINE:
! compute knowclass variable C whose values represent all
! combinations of values of X and K
  IF K == 0 AND X == 0 THEN C = 1 ;
  IF K == 1 AND X == 0 THEN C = 2 ;
  IF K == 0 AND X == 1 THEN C = 3 ;
  IF K == 1 AND X == 1 THEN C = 4 ;
  IF K == 0 AND X == 2 THEN C = 5 ;
  IF K == 1 AND X == 2 THEN C = 6 ;
VARIABLE: NAMES = K X y32 y22 y12 eta2 y31 y21 y11 eta1 Ix2 Ix1;
  USEVARIABLES = y32 y22 y12 y31 y21 y11 C;
  MISSING =.;
  CLASSES = cell( 6 ); ! 6 cells (3x2 combinations of X and K)
  KNOWNCLASS = cell( C=1 C=2 C=3 C=4 C=5 C=6 );
ANALYSIS: TYPE = MIXTURE;
```

Example 3

M*plus* syntax for example 3 (continued):

```
MODEL:
%OVERALL%
 ! measurement model for latent variables
 eta2 by y12@1 y22@1 y32@1;
 eta1 by y11@1 y21@1 y31@1;
 [y11@0 y21@0 y31@0 y12@0 y22@0 y32@0];
 [eta2* eta1*];

! regression of eta2 on eta1
 eta2 ON eta1;

! coefficients for group sizes 1-5 (logistic metric)
! group 6 is the reference group
 [cell#1] (LC1);
 [cell#2] (LC2);
 [cell#3] (LC3);
 [cell#4] (LC4);
 [cell#5] (LC5);
```

M*plus* syntax for example 3 (continued):

```
! regression of eta2 on eta1 in each cell
%cell#1%
eta2 ON eta1(a0_0_1); ! regression coefficient
[eta1] (EzC1); ! cell specific mean of eta1
[eta2] (a0_0_0); ! intercept of regression

%cell#2%
eta2 ON eta1(a0_1_1);
[eta1] (EzC2);
[eta2] (a0_1_0);

%cell#3%
eta2 ON eta1(a1_0_1);
[eta1] (EzC3);
[eta2] (a1_0_0);
```

Example 3

M*plus* syntax for example 3 (continued):

```
%cell#4%
eta2 ON eta1(a1_1_1);
[eta1] (EzC4);
[eta2] (a1_1_0);

%cell#5%
eta2 ON eta1(a2_0_1);
[eta1] (EzC5);
[eta2] (a2_0_0);

%cell#6%
eta2 ON eta1(a2_1_1);
[eta1] (EzC6);
[eta2] (a2_1_0);
```

M*plus* syntax for example 3 (continued):

```
MODEL CONSTRAINT:

NEW(PC1 PC2 PC3 PC4 PC5 PC6 PK1 ! P(C=c) and P(K=1)
Ez ! unconditional mean eta1
EzK1 ! mean of K*eta1 interaction term
b0_0_0 b0_0_1 b0_1_0 b0_1_1
b1_0_0 b1_0_1 b1_1_0 b1_1_1
b2_0_0 b2_0_1 b2_1_0 b2_1_1
g1_0_0 g1_0_1 g1_1_0 g1_1_1
g2_0_0 g2_0_1 g2_1_0 g2_1_1
Eg1 Eg2);

! cell probabilities
PC1 = EXP(LC1) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
PC2 = EXP(LC2) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
PC3 = EXP(LC3) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
PC4 = EXP(LC4) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
PC5 = EXP(LC5) / (EXP(LC1) +EXP(LC2) +EXP(LC3) +EXP(LC4) +EXP(LC5)
PC6 = 1 - (PC1 +PC2 +PC3 +PC4 +PC5);
```

Example 3

M*plus* syntax for example 3 (continued):

```
PK1 = PC2 +PC4 +PC6;
Ez = EzC1*PC1 +EzC2*PC2 +EzC3*PC3 +EzC4*PC4 +EzC5*PC5 +EzC6*PC6;
EzK1 = EzC2*PC2 + EzC4*PC4 + EzC6*PC6; ! mean of K*eta1 interactio

! coefficients of (X=x)-conditional regressions of eta2 on K and e
b0_0_0 = a0_0_0;
b0_0_1 = a0_0_1;
b0_1_0 = a0_1_0 - a0_0_0;
b0_1_1 = a0_1_1 - a0_0_1;
b1_0_0 = a1_0_0;
b1_0_1 = a1_0_1;
b1_1_0 = a1_1_0 - a1_0_0;
b1_1_1 = a1_1_1 - a1_0_1;
b2_0_0 = a2_0_0;
b2_0_1 = a2_0_1;
b2_1_0 = a2_1_0 - a2_0_0;
b2_1_1 = a2_1_1 - a2_0_1;
```

Example 3

M*plus* syntax for example 3 (continued):

```
! coefficients of effect functions g1 and g2
g1_0_0 = b1_0_0 - b0_0_0;
g1_0_1 = b1_0_1 - b0_0_1;
g1_1_0 = b1_1_0 - b0_1_0;
g1_1_1 = b1_1_1 - b0_1_1;

g2_0_0 = b2_0_0 - b0_0_0;
g2_0_1 = b2_0_1 - b0_0_1;
g2_1_0 = b2_1_0 - b0_1_0;
g2_1_1 = b2_1_1 - b0_1_1;

! average effects
Eg1 = g1_0_0 + g1_0_1*Ez + g1_1_0*PK1 + g1_1_1*EzK1;
Eg2 = g2_0_0 + g2_0_1*Ez + g2_1_0*PK1 + g2_1_1*EzK1;
```

Software

Software

- M*plus* can be used for the models presented here.
- Since specification of the MODEL CONSTRAINT can be tedious for complex models, we are developing an R package that does this automatically for you. The R package has not been published yet, but can be obtained by emailing Lisa (`Lisa.Dietzfelbinger@uni-jena.de`) or me (`Axel.Mayer@ugent.be`).
- For those of you who also use other software, we developed a lavaan based version of the program, which includes a graphical user interface and can be obtained via my Github repository (`https://github.com/amayer2010/effectliter`).

Thank you for your attention.