**Supplementary Materials**

for

**The fixed versus random effects debate and how it relates to centering in multilevel modeling**

E. L. Hamaker & B. Muthén

R-code for simulating (and analyzing) data and Mplus input for analyzing data……………p. 2

R-code for simulating data for 1000 replications and Mplus input for analyzing data …………p. 13

Empirical application……………p. 25

**R-code for simulating (and analyzing) a single data set**

**and Mplus input for analyzing a single data set**

**R-code**

######################################

#### SIMULATE DATA WITH DIFFERENT ####

###### WITHIN AND BETWEEN SLOPES #####

######################################

rm(list=ls())

library(lme4)

set.seed(3859)

N<-5000 # number of clusters

n<-4 # number of observations within a cluster

Y <- matrix(,N,n) # to store the outcome

X <- matrix(,N,n) # to store the predictor

X.TrueCent <- matrix(,N,n) # to store the true centered predictor

########################

#### SET PARAMETER VALUES ####

########################

# PREDICTOR

sdX.within <- sqrt(1) # within-person variance

sdX.between <- sqrt(4) # between-person variance

# INTERCEPT LEVEL 2

g.00 <- 0 # Grand intercept

g.01 <- 2 # between-cluster slope

sd.u0 <- 1 # SD of residuals intercept at level 2

# SLOPE LEVEL 2

g.10 <- 1 # fixed within-cluster slope

sd.u1 <- 0 # SD of within-cluster slope at level 2

# RESIDUALS AT LEVEL 1

sd.e <- 1 # residual SD at level 1

#######################

#### SIMULATE DATA ####

#######################

for (j in 1:N)

{ # sample mean on x in cluster j

X.mean.j <- rnorm(1,mean=0,sd=sdX.between)

# sample x in cluster j

X.j <- rnorm(n,mean=X.mean.j,sd=sdX.within)

X[j,] <- X.j

# sample slope in cluster j (here identical across individuals)

b1.j <- rnorm(1,g.10,sd.u1)

# sample intercept in cluster j (level 2 expression)

b0.j <- rnorm(1,g.00,sd.u0) + g.01\*X.mean.j

# sample y (level 1 expression)

Y[j,] <- b0.j + b1.j\*(X.j-X.mean.j) + rnorm(n,0,sd.e)

}

########################

### DATA PREPARATION ###

########################

Y <- c(t(Y))

X <- c(t(X))

Cluster <- rep(1:N, each=n)

Time <- rep(1:n,N)

#####################################

### ESTIMATE SAMPLE CLUSTER MEANS ###

###### AND CENTER X WITH THIS #######

#####################################

cluster.means <- NA

cluster.j <- 0

for (j in 1:N){

cluster.mean <- rep(mean(X[((cluster.j\*n)+1):((cluster.j\*n)+n)]),n)

cluster.means <- c(cluster.means,cluster.mean)

cluster.j <- cluster.j + 1

}

cluster.means <- as.vector(cluster.means)[-1]

X.cent <- X - cluster.means

data.cent <- as.data.frame(cbind(Y, X.cent,X, Cluster, cluster.means))

########################

### ESTIMATE MODELS ####

########################

# L1: RAW X (RE MODEL)

summary(lmer(Y ~ X + (1 |Cluster), data = data.cent, REML = F))

# L2: CENTERED X (FE MODEL)

summary(lmer(Y ~ X.cent + (1 |Cluster), data = data.cent, REML = F))

# L3a: CENTERED X PLUS SAMPLE MEANS ON X (WITHIN-BETWEEN MODEL)

summary(lmer(Y ~ X.cent + cluster.means + (1 |Cluster), data = data.cent, REML = F))

# L4: RAW X PLUS SAMPLE MEANS ON X (MUNDLAK'S CONTEXTUAL MODEL)

summary(lmer(Y ~ X + cluster.means + (1 |Cluster), data = data.cent, REML = F))

##################

### WRITE DATA ###

##################

write.table(cbind(round(data.cent,5),Time),

file="BWdataT4.dat",col.names=FALSE,row.names=FALSE)

**Method L1**

TITLE: Multilevel model

DATA: FILE = BWdataT4.dat;

VARIABLE: NAMES = y xcent x clus clusmean;

USEVARIABLE = y x;

WITHIN = x;

BETWEEN = ;

CLUSTER = clus;

ANALYSIS: TYPE = TWOLEVEL;

MODEL:

%WITHIN%

y ON x;

%BETWEEN%

y;

**Mplus code method L2: Multilevel model with within-person centered predictor**

DATA: FILE = BWdataT4.dat;

VARIABLE: NAMES = y xcent x clus clusmean;

USEVARIABLE = y xcent;

WITHIN = xcent;

BETWEEN =;

CLUSTER = clus;

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

y ON xcent;

%BETWEEN%

y;

**Or center the predictor within Mplus with the sample mean per cluster (using the DEFINE statement):**

DATA: FILE = BWdataT4.dat;

VARIABLE: NAMES = y xcent x clus clusmean;

USEVARIABLE = y x;

WITHIN = x;

BETWEEN = ;

CLUSTER = clus;

DEFINE: CENTER x (GROUPMEAN);

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

y ON x;

%BETWEEN%

y;

**Mplus code for method L3a: Multilevel model with centered X variable and sample mean as a between level predictor:**

DATA: FILE = BWdataT4.dat;

VARIABLE: NAMES = y xcent x clus clusmean;

USEVARIABLE = y xcent clusmean;

WITHIN = xcent;

BETWEEN = clusmean;

CLUSTER = clus;

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

y ON xcent;

%BETWEEN%

y ON clusmean;

**Or center the predictor within Mplus with the sample mean per cluster, and save these cluster means (meanx) to use as a between level predictor (using the DEFINE statement):**

DATA: FILE = BWdataT4.dat;

VARIABLE: NAMES = y xcent x clus clusmean;

USEVARIABLE = y x meanx;

WITHIN = x;

BETWEEN = meanx;

CLUSTER = clus;

DEFINE: CENTER x (GROUPMEAN);

meanx = CLUSTER\_MEAN (x);

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

y ON x;

%BETWEEN%

y ON meanx;

**Mplus code for method L3b: Multilevel model with centered X variable using latent mean centering**

DATA: FILE = BWdataT4.dat;

VARIABLE: NAMES = y xcent x clus clusmean;

USEVARIABLE = y x;

WITHIN = ;

BETWEEN = ;

CLUSTER = clus;

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

y ON x;

%BETWEEN%

y ON x;

**Mplus code for method L4: Multilevel model with raw X variable and cluster means (clusmean) as between level predictor:**

DATA: FILE = BWdataT4.dat;

VARIABLE: NAMES = y xcent x clus clusmean;

USEVARIABLE = y x clusmean;

WITHIN = x;

BETWEEN = clusmean;

CLUSTER = clus;

ANALYSIS: TYPE = TWOLEVEL;

MODEL:

%WITHIN%

y ON x;

%BETWEEN%

y ON clusmean;

**Or computing the cluster mean (meanx) inside Mplus using the DEFINE statement:**

DATA: FILE = BWdataT4.dat;

VARIABLE: NAMES = y xcent x clus clusmean;

USEVARIABLE = y x meanx;

WITHIN = x;

BETWEEN = meanx;

CLUSTER = clus;

DEFINE: meanx = CLUSTER\_MEAN (x);

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

y ON x;

%BETWEEN%

y ON meanx;

**Mplus code for method W1: Model with a random intercept that is uncorrelated with X (the RE model)**

DATA: FILE = BWdataT4.dat;

DATA LONGTOWIDE: ! Convert the long-format data file to wide-format

LONG = y | x;

WIDE = y1-y4 | x1-x4;

IDVARIABLE = clus;

REPETITION = Time (1-4);

VARIABLE: NAMES = y xcent x clus clusmean time;

USEVARIABLE = y1-y4 x1-x4;

MODEL: alpha by y1-y4@1; ! random effect

y1-y4 PON x1-x4 (b1);

alpha WITH x1-x4@0; ! fixing the correlations to zero

! Stationarity constraints (optional):

! this allows for comparing it to the long format analyses

x1-x4 WITH x1-x4 (r);

x1-x4 (vx);

[y1-y4] (b0);

[x1-x4] (mu);

y1-y4 (v);

**Mplus code for method W2: Model with a random intercept that is correlated with X (the FE model)**

DATA: FILE = BWdataT4.dat;

DATA LONGTOWIDE: ! Convert the long-format data file to wide-format

LONG = y | x;

WIDE = y1-y4 | x1-x4;

IDVARIABLE = clus;

REPETITION = Time (1-4);

VARIABLE: NAMES = y xcent x clus clusmean time;

USEVARIABLE = y1-y4 x1-x4;

MODEL: alpha by y1-y4@1; ! random effect

y1-y4 PON x1-x4 (b1);

! Stationarity constraints (optional):

! this allows for comparing it to the long format analyses

x1-x4 WITH x1-x4 (r);

x1-x4 (vx);

[y1-y4] (b0);

[x1-x4] (mu);

y1-y4 (v);

**Mplus code for method W3: Model with a random intercept that is correlated with X (the FE model)**

DATA: FILE = BWdataT4.dat;

DATA LONGTOWIDE: ! Convert the long-format data file to wide-format

LONG = y | x;

WIDE = y1-y4 | x1-x4;

IDVARIABLE = clus;

REPETITION = Time (1-4);

VARIABLE: NAMES = y xcent x clus clusmean time;

USEVARIABLE = y1-y4 x1-x4;

MODEL: alpha by y1-y4@1; ! person-specific factor for y

eta BY x1-x4@1; ! person-specific factor for x

alpha ON eta; ! between-person regression

! Create within-person centered X variables

cx1 BY x1; cx2 BY x2; cx3 BY x3; cx4 BY x4;

! Overrule Mplus defaults

x1-x4@0; ! measurement error variances fix to zero

eta cx1-cx4 WITH cx1-cx4@0; ! default covariances fixed to zero

y1-y4 PON cx1-cx4 (b1); ! within-person regression

! Stationarity constraints (optional)

! this allows for comparing it to the long format analyses

[y1-y4] (b0); ! stationarity constraints on intercepts of y

[x1-x4] (mu); ! stationarity constraints on means of x

y1-y4 (v); ! stationarity constraints on residual variances of y

cx1-cx4 (cvx); ! stationarity constraints on variances of within-person part of x

**Mplus code for method W4: Regressing the outcome on the raw X, while including the within-person mean on X as a predictor of the random intercept**

DATA: FILE = BWdataT4.dat;

DATA LONGTOWIDE:

LONG = y | x;

WIDE = y1-y4 | x1-x4;

IDVARIABLE = clus;

REPETITION = Time (1-4);

VARIABLE: NAMES = y xcent x clus clusmean time;

USEVARIABLE = y1-y4 x1-x4;

MODEL: alpha by y1-y4@1; ! person-specific factor for y

eta BY x1-x4@1; ! person-specific factor for x

alpha ON eta; ! between-person regression

y1-y4 PON x1-x4 (b1); ! within-person regression

! Stationarity constraints (optional)

! this allows for comparing it to the long format analyses

[y1-y4] (b0); ! stationarity constraints on intercepts of y

[x1-x4] (mu); ! stationarity constraints on means of x

y1-y4 (v); ! stationarity constraints on residual variances of y

x1-x4 (cvx); ! stationarity constraints on residual variances of x

**R-code for simulating data for 1000 replications**

**and Mplus input to analyze these data**

######################################

#### SIMULATE DATA WITH DIFFERENT ####

###### WITHIN AND BETWEEN SLOPES #####

######################################

rm(list=ls())

set.seed(2328)

R <- 1000 # numbser of replications

N <- 100 # number of clusters

N <- 4 # number of observations within a cluster

Lnames <- NA

Wnames <- NA

##############################

#### SET PARAMETER VALUES ####

##############################

# PREDICTOR

sdX.within <- sqrt(1) # within-person variance

sdX.between <- sqrt(4) # between-person variance

# INTERCEPT LEVEL 2

g.00 <- 0 # Grand intercept

g.01 <- .5 # between-cluster slope

sd.u0 <- 2 # SD of residuals intercept at level 2

# SLOPE LEVEL 2

g.10 <- 1 # fixed within-cluster slope

sd.u1 <- 0 # SD of within-cluster slope at level 2

# RESIDUALS AT LEVEL 1

sd.e <- 2 # residual SD at level 1

v.y.within <- g.10^2\*sdX.within^2 + sd.e^2

v.y.between <- g.01^2\*sdX.between^2 + sd.u0^2

v.y.between/(v.y.between+v.y.within)

#######################

#### SIMULATE DATA ####

#######################

for (r in 1:R){

Y <- matrix(,N,n) # to store the outcome

X <- matrix(,N,n) # to store the predictor

for (j in 1:N)

{ # sample mean on x in cluster j

X.mean.j <- rnorm(1,mean=0,sd=sdX.between)

# sample x in cluster j

X.j <- rnorm(n,mean=X.mean.j,sd=sdX.within)

# sample slope in cluster j (here identical across individuals)

b1.j <- rnorm(1,g.10,sd.u1)

# sample intercept in cluster j (level 2 expression)

b0.j <- rnorm(1,g.00,sd.u0) + g.01\*X.mean.j

# sample y (level 1 expression)

Y[j,] <- b0.j + b1.j\*(X.j-X.mean.j) + rnorm(n,0,sd.e)

# save x and the within-person centered

# x using the true mean on x

X[j,] <- X.j

}

dataW <- cbind(Y,X)

###################################

### CONVERT DATA TO LONG FORMAT ###

###################################

YL <- c(t(Y))

XL <- c(t(X))

Cluster <- rep(1:N, each=n)

Occasion <- rep(1:n,N)

dataL <- as.data.frame(cbind(YL, XL, Cluster, Occasion))

##################

### WRITE DATA ###

##################

write.table(round(dataW,5),file=paste0("N100T4Wrep",

r,".dat"),col.names=FALSE,row.names=FALSE)

write.table(round(dataL,5),file=paste0("N100T4Lrep",

r,".dat"),col.names=FALSE,row.names=FALSE)

Lnames <- c(Lnames,paste0("N100T4Lrep",r,".dat"))

Wnames <- c(Wnames,paste0("N100T4Wrep",r,".dat"))

}

##############################################

### SAVE DATA IN LONG AND WIDE FORMAT ########

##############################################

write.table(Lnames[-1],quote=FALSE,file="N100T4Lreplist.dat",

col.names=FALSE,row.names=FALSE)

write.table(Wnames[-1],quote=FALSE,file="N100T4Wreplist.dat",

col.names=FALSE,row.names=FALSE)

**Mplus code to compute the intraclass correlation**

TITLE: Simulation study based on 1000 replications

Data in long format

DATA: FILE = N100T4Lreplist.dat;  ! Panel data with T=4

! FILE = N100T40Lreplist.dat;  ! Intensive longitudinal data with T=40

TYPE = MONTECARLO;

VARIABLE: NAMES = y x clus;

USEVARIABLE = y x;

WITHIN = ;

BETWEEN = ;

CLUSTER = clus;

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

y (vyw);

x (vxw);

%BETWEEN%

y (vyb);

x (vxb);

MODEL CONSTRAINT:

new (iccY); iccY=vyb/(vyb+vyw);

new (iccX); iccX=vxb/(vxb+vxw);

OUTPUT: TECH1;

**Mplus code method L1**

TITLE: Simulation study based on 1000 replications

Data in long format

DATA: FILE = N100T4Lreplist.dat;  ! Panel data with T=4

! FILE = N100T40Lreplist.dat;  ! Intensive longitudinal data with T=40

TYPE = MONTECARLO;

VARIABLE: NAMES = y x clus;

USEVARIABLE = y x;

WITHIN = x;

BETWEEN = ;

CLUSTER = clus;

ANALYSIS: TYPE = TWOLEVEL;

MODEL:

%WITHIN%

y ON x\*1;

%BETWEEN%

y\*1;

[y\*0];

OUTPUT: TECH1;

**Mplus code method L2**

TITLE: Simulation study based on 1000 replications

Data in long format

DATA: FILE = N100T4Lreplist.dat;  ! Panel data with T=4

! FILE = N100T40Lreplist.dat;  ! Intensive longitudinal data with T=40

TYPE = MONTECARLO;

VARIABLE: NAMES = y x clus;

USEVARIABLE = y x;

WITHIN = x;

BETWEEN = ;

CLUSTER = clus;

DEFINE: CENTER x (GROUPMEAN);

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

y ON x\*1;

x\*1;

y\*1;

%BETWEEN%

y\*1;

[y\*0];

OUTPUT: TECH1;

**Mplus code method L3a**

TITLE: Simulation study based on 1000 replications

Data in long format

DATA: FILE = N100T4Lreplist.dat;  ! Panel data with T=4

! FILE = N100T40Lreplist.dat;  ! Intensive longitudinal data with T=40

TYPE = MONTECARLO;

VARIABLE: NAMES = y x clus;

USEVARIABLE = y x meanX;

WITHIN = x;

BETWEEN = meanX;

CLUSTER = clus;

DEFINE: meanX = CLUSTER\_MEAN (x);

CENTER x (GROUPMEAN);

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

y ON x\*1;

x\*1;

y\*4;

%BETWEEN%

y ON meanX\*.5;

meanX\*4;

y\*4;

[meanX\*0];

[y\*0];

OUTPUT: TECH1;

**Mplus code method L3b**

TITLE: Simulation study based on 1000 replications

Data in long format

DATA: FILE = N100T4Lreplist.dat;  ! Panel data with T=4

! FILE = N100T40Lreplist.dat;  ! Intensive longitudinal data with T=40

TYPE = MONTECARLO;

VARIABLE: NAMES = y x clus;

USEVARIABLE = y x;

WITHIN = ;

BETWEEN = ;

CLUSTER = clus;

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

y ON x\*1;

x\*1;

y\*4;

%BETWEEN%

y ON x\*.5;

x\*4;

y\*4;

[x\*0];

[y\*0];

OUTPUT: TECH1;

**Mplus code method L4**

TITLE: Simulation study based on 1000 replications

Data in long format

DATA: FILE = N100T4Lreplist.dat;  ! Panel data with T=4

! FILE = N100T40Lreplist.dat;  ! Intensive longitudinal data with T=40

TYPE = MONTECARLO;

VARIABLE: NAMES = y x clus;

USEVARIABLE = y x meanX;

WITHIN = x;

BETWEEN = meanX;

CLUSTER = clus;

DEFINE: meanX = CLUSTER\_MEAN (x);

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

y ON x\*1;

x\*1;

y\*4;

%BETWEEN%

y ON meanX\*-.5;

meanX\*4;

y\*4;

[meanX\*0];

[y\*0];

OUTPUT: TECH1;

**Mplus code method W1**

TITLE: Simulation study based on 1000 replications

Data in wide format

DATA: FILE = N100T4Wreplist.dat;

TYPE = MONTECARLO;

VARIABLE: NAMES = y1-y4 x1-x4;

MODEL: alpha by y1-y4@1; ! random effect

y1-y4 PON x1-x4\*1 (b1);

alpha WITH x1-x4@0; ! fixing the correlations to zero

! Stationarity constraints (optional)

x1-x4 WITH x1-x4 (r);

x1-x4 (vx);

[y1-y4] (b0);

[x1-x4] (mu);

y1-y4 (v);

OUTPUT: TECH1;

**Mplus code method W2**

TITLE: Simulation study based on 1000 replications

Data in wide format

DATA: FILE = N100T4Wreplist.dat;

TYPE = MONTECARLO;

VARIABLE: NAMES = y1-y4 x1-x4;

MODEL: alpha by y1-y4@1; ! Latent variable

y1-y4 PON x1-x4\*1 (b1);

! Stationarity constraints

[y1-y4] (b0);

[x1-x4] (mu);

x1-x4 (vx);

y1-y4 (v);

x1-x4 WITH x1-x4 (r);

x1-x4 WITH alpha (k);

OUTPUT: TECH1;

**Mplus code method W3**

TITLE: Simulation study based on 1000 replications

Data in wide format

DATA: FILE = N100T4Wreplist.dat;

TYPE = MONTECARLO;

VARIABLE: NAMES = y1-y4 x1-x4;

MODEL: alpha by y1-y4@1; ! person-specific factor for y

eta BY x1-x4@1; ! person-specific factor for x

alpha ON eta\*.5; ! between-person regression

alpha\*4; ! between-person residual variance on y

eta\*4; ! between-person variance on x

! Create within-person centered X variables

cx1 BY x1;

cx2 BY x2;

cx3 BY x3;

cx4 BY x4;

x1-x4@0; ! measurement error variances fix to zero

eta cx1-cx4 WITH cx1-cx4@0; ! default covariances fixed to zero

y1-y4 PON cx1-cx4\*1 (b1); ! within-person regression

! Stationarity constraints (optional)

[y1-y4\*0] (b0); ! stationarity constraints on intercepts of y

[x1-x4\*0] (mu); ! stationarity constraints on means of x

y1-y4\*4 (v); ! stationarity constraints on residual variances of y

cx1-cx4\*1 (cvx); ! stationarity constraints on variances of within part of x

OUTPUT: TECH1;

**Mplus code method W4**

TITLE: Simulation study based on 1000 replications

Data in wide format

DATA: FILE = N100T4Wreplist.dat;

TYPE = MONTECARLO;

VARIABLE: NAMES = y1-y4 x1-x4;

MODEL: alpha by y1-y4@1; ! person-specific factor for y

eta BY x1-x4@1; ! person-specific factor for x

alpha ON eta\*-0.5; ! contextual effect (b^(b) - b^(w))

y1-y4 PON x1-x4\*1 (b1); ! within-person regression

! Stationarity constraints (optional)

! this allows for comparing it to the long format analyses

[y1-y4] (b0); ! stationarity constraints on intercepts of y

[x1-x4] (mu); ! stationarity constraints on means of x

y1-y4 (v); ! stationarity constraints on residual variances of y

x1-x4 (cvx); ! stationarity constraints on residual variances of x

OUTPUT: TECH1;

**Empirical application**

The data here come from Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, . . . Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. PLoS ONE, 8 , e60188, 1-13. doi: 10.1371/journal.pone.0060188; and were obtained here: <http://journals.plos.org/plosone/article/file?type=supplementary&id=info:doi/10.1371/journal.pone.0060188.s004>

The analyses are based on the first phase of that study only (prior to intervention); the dataset Bringmann1.dat only contains data from the first phase, from which we make use of momentary somberness (outcome) and event (time-varying covariate). Additionally, we use the level 2 variable neuroticism, which was measured prior to the intensive longitudinal measurements (time-invariant covariate). This variable was rescaled in Mplus prior to analysis (using the DEFINE statement, see below), for computational purposes.

**Mplus input:**

TITLE: Multilevel model with latent mean centering for the X variable and a fixed slope

DATA: FILE = Bringmann1.dat;

VARIABLE: NAMES = ID dayno beepno informat04 st\_period

cheerful event worry fearful somber

relaxed neurotic;

USEVARIABLE = somber event neurotic N;

CLUSTER = ID;

BETWEEN = neurotic N;

MISSING = ALL(-99);

DEFINE: N = neurotic/10;

ANALYSIS: TYPE = TWOLEVEL;

MODEL: %WITHIN%

somber ON event;

%BETWEEN%

somber ON event N;

event ON N;

MODEL INDIRECT: somber IND N;

OUTPUT: TECH1 STDYX SAMPSTAT;