

VERSION 7.2

Mplus LANGUAGE ADDENDUM

This addendum describes changes introduced in Version 7.2. They include corrections to minor problems that have been found since the release of Version 7.11 in June 2013 as well as the following new features:

- Mixture modeling with non-normal distributions
- Mediation analysis with direct and indirect effects based on counterfactuals (causal inference)
- Latent class and latent transition analysis with residual covariances
- Restructured routines for continuous-time survival analysis with latent variables
- Structural equation modeling (SEM) with non-normal distributions
- New order of operations for the DEFINE command
- Double do loops for the DEFINE, MODEL CONSTRAINT, MODEL TEST, and MODEL PRIORS commands
- ALIGNMENT option for binary outcomes and TYPE=COMPLEX using MLR
- Bootstrap standard errors and confidence intervals for maximum likelihood estimation with ALGORITHM=INTEGRATION
- Standard errors for TECH4 for the Delta parameterization of weighted least squares and z-tests and p-values for TECH4
- Standardized coefficients with standard errors for models with covariates using weighted least squares estimation
- New plots: Estimated distributions; Estimated medians, modes, and percentiles; and Scatterplots of residuals

- For Monte Carlo studies, TYPE=TWOLEVEL, and ESTIMATOR=BAYES, the output contains a table showing the correlation and mean square error comparing true and estimated factor scores
- For the ALIGNMENT option and real data, RANKING=filename.csv; in the SAVEDATA command produces a comma-delimited file that shows the rankings of groups based on the group factor means and also shows the significance of the factor mean differences
- For Monte Carlo studies using the ALIGNMENT option, the output contains a table showing the correlation and mean square error comparing true and estimated factor means
- Parameter names given for parameter numbers listed as non-identified
- Several lines can be commented out by starting the first line with !* and ending the last line with *!
- New features for the Mac Editor: Correction to undo function and addition of All file types

MIXTURE MODELING WITH NON-NORMAL DISTRIBUTIONS

Mixture modeling with non-normal distributions is available using the DISTRIBUTION option of the ANALYSIS command in conjunction with TYPE=MIXTURE. The DISTRIBUTION option has four settings: NORMAL, SKEWNORMAL, TDISTRIBUTION, and SKEWT. The default is NORMAL. The DISTRIBUTION option can be used with only continuous variables although the analysis model can contain other types of variables. The DISTRIBUTION option cannot be used with models that require numerical integration.

The SKEWNORMAL and SKEWT settings have a special skew parameter for each observed and latent variable that is related to

the skewness of the variable. It is specified by mentioning the name of the variable in squiggly brackets. For example, the skew parameter for a variable y is specified as $\{y\}$. Skew parameters are free and unequal across classes with starting values of one. They can be constrained to be equal or fixed at a particular value. The TDISTRIBUTION and SKEWT settings have a special degree of freedom parameter for each class that is related to the degrees of freedom in the t-distribution. The degree of freedom parameter is specified by putting df in squiggly brackets, for example, $\{df\}$. The degree of freedom parameter is free and unequal across classes with a starting value of one. It can be constrained to be equal or fixed at a particular value. The SKEWNORMAL setting can capture skewness less than the absolute value of one, whereas the SKEWT setting has no such limitations. Asparouhov and Muthén (2014a) describes the theory behind the implementation of mixture modeling with non-normal distributions.

MEDIATION ANALYSIS WITH EFFECTS BASED ON COUNTERFACTUALS (CAUSAL INFERENCE)

Causally-defined direct and indirect effects in mediation analysis as described in Muthén (2011) and Muthén and Asparouhov (2014) are available using MODEL INDIRECT with maximum likelihood estimation. The effects are available for a single mediator and a single outcome. The causally-defined direct and indirect effects are different than the usual direct and indirect effects of SEM in several cases. Cases like this that are covered in MODEL INDIRECT include models with a binary outcome, a count outcome, a binary mediator, and moderation that involves the mediator. The observed exogenous variable in these direct and indirect effects is referred to in the causal-effects literature as the treatment variable or exposure variable. We refer to it as the exposure variable. It can be binary or continuous. The IND and MOD options of MODEL INDIRECT are used to specify the causally-defined direct and indirect effects.

The IND option is used to specify a specific indirect effect when there is no moderation. Following is an example of how to specify the IND option with a binary exposure variable:

MODEL INDIRECT:
y IND m x;

where y is the outcome, m is the mediator, and x is a binary exposure variable. The outcome and mediator can be continuous latent variables.

When the exposure variable is continuous, two values must be given in parentheses following the exposure variable. The causal effects are computed comparing these two values. The default is one for the first value and zero for the second value corresponding to a binary exposure variable such as comparing a treatment group to a control group. Following is an example of how to specify the IND option with a continuous exposure variable:

MODEL INDIRECT:
y IND m x (1 -1);

where y is the outcome, m is the mediator, and x is the continuous exposure variable. The two values in parentheses following x are the values used to compute the causal effects, in this case, comparing 1 to -1, which represents an increase in x from -1 to 1. The outcome, mediator, and the exposure variable can be continuous latent variables.

The MOD option is used to specify a specific indirect effect when there is moderation. The MOD option can have three, four, or five arguments after MOD. The MOD option has three arguments when there is an interaction between the exposure variable and the

mediator. Following is an example of how to specify the MOD option with three arguments:

MODEL INDIRECT:
y MOD m xm x;

where y is the outcome, m is the mediator, xm is the interaction between x and m, and x is a binary exposure variable. The variables must be given in this order. The outcome can be a continuous latent variable.

When a model contains an exogenous moderator, in addition to the regular results, a plot is available showing the effects and their confidence intervals as a function of different values of the exogenous moderator. The MOD option with an exogenous moderator can have four or five arguments.

The MOD option followed by four arguments has two specifications. The exogenous moderator can interact with either the mediator or the exposure variable. Following is an example of how to specify the MOD option with four arguments when the exogenous moderator interacts with the mediator:

MODEL INDIRECT:
y MOD m z (-1 1 0.1) mz x;

where y is the outcome, m is the mediator, z is the exogenous moderator, mz is the interaction between m and z, and x is a binary exposure variable. The variables must be given in this order. The numbers in parentheses following z are the lower limit, upper limit, and the increment to be used in evaluating and plotting the direct and indirect effects. The outcome can be a continuous latent variable.

Following is an example of how to specify the MOD option with four arguments when the exogenous moderator interacts with the mediator and the exposure variable is continuous:

MODEL INDIRECT:

y MOD m z (-1 1 0.1) mz x (1 -1);

where y is the outcome, m is the mediator, z is the exogenous moderator, mz is the interaction between m and z, and x is a continuous exposure variable. The variables must be given in this order. The numbers in parentheses following z are the lower limit, upper limit, and the increment to be used in evaluating and plotting the direct and indirect effects. The two values in parentheses following x are the values used to compute the causal effects, in this case, comparing 1 to -1, which represents an increase in x from -1 to 1. The outcome and the exposure variable can be continuous latent variables.

Following is an example of how to specify the MOD option with four arguments when the exogenous moderator interacts with the exposure variable:

MODEL INDIRECT:

y MOD m z (-1 1 0.1) xz x;

where y is the outcome, m is the mediator, z is the exogenous moderator, xz is the interaction between x and z, and x is a binary exposure variable. The variables must be given in this order. The numbers in parentheses following z are the lower limit, upper limit, and the increment to be used in evaluating and plotting the direct and indirect effects. The outcome can be a continuous latent variable.

The MOD option has five arguments when an exogenous moderator interacts with both the mediator and the exposure

variable. Following is an example of how to specify the MOD option with five arguments:

```
MODEL INDIRECT:  
y MOD m z (-1 1 0.1) mz xz x;
```

where y is the outcome, m is the mediator, z is the exogenous moderator, mz is the interaction between m and z, xz is the interaction between x and z, and x is a binary exposure variable. The variables must be given in this order. The numbers in parentheses following z are the lower limit, upper limit, and the increment to be used in evaluating and plotting the direct and indirect effects. The outcome can be a continuous latent variable.

If the special direct effect referred to as the controlled direct effect is wanted, a value of the mediator must be given in parentheses following the mediator variable. Following is an example of how to specify the IND option with a binary exposure variable:

```
MODEL INDIRECT:  
y IND m (2) x;
```

where y is the outcome, m is the mediator, and x is the binary exposure variable. The value in parentheses following m is used to compute the controlled direct effect.

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

TITLE:	this is an example of moderated mediation with a plot of the indirect effect as in
--------	---

```

                                Example 3.18 but using bootstrap and
                                maximum likelihood estimation
DATA:      FILE = ex3.18.dat;
VARIABLE:  NAMES = y m x z;
           USEVARIABLES = y m x z xz;
DEFINE:    xz = x*z;
ANALYSIS:  BOOTSTRAP = 500;
MODEL:     y ON m xz z;
           m ON z xz x;
MODEL INDIRECT:
           y MOD m z (-2 2 0.1) xz x;
PLOT:      TYPE = PLOT2;
OUTPUT:    CINTERVAL (BCBOOTSTRAP);

```

LATENT CLASS AND LATENT TRANSITION ANALYSIS WITH RESIDUAL COVARIANCES

For latent class analysis and latent transition analysis, `PARAMETERIZATION=RESCOVARIANCES` allows the `WITH` option to be used to specify residual covariances for binary and ordered categorical (ordinal) outcomes using maximum likelihood estimation (Asparouhov & Muthén, 2014b). These residual covariances can be free across classes, constrained to be equal across classes, or appear in only certain classes. Following is a partial input for a latent class analysis where the residual covariances are held equal across classes:

```

VARIABLE:
CATEGORICAL = u1-u4;
CLASSES = c(2);
ANALYSIS:
TYPE=MIXTURE;
PARAMETERIZATION=RESCOVARIANCES;
MODEL:
%OVERALL%
u1 WITH u3;

```

Following is a partial input for a latent class analysis where the residual covariances are not held equal across classes:


```
VARIABLE:  
CATEGORICAL = u1-u4;  
CLASSES = c(2);  
ANALYSIS:  
TYPE=MIXTURE;  
PARAMETERIZATION=RESCOV;  
MODEL:  
%OVERALL%  
u1 WITH u3;  
%c#1%  
u1 WITH u3;
```

Following is a partial input for a latent transition analysis where the residual covariances are allowed in only specific classes:

```
VARIABLE:  
CATEGORICAL = u1-u8;  
CLASSES = c1 (3) c2 (3);  
ANALYSIS:  
TYPE=MIXTURE;  
PARAMETERIZATION=RESCOV;  
MODEL:  
%OVERALL%  
c2 ON c1;  
%c1#2.c2#2%  
u1 WITH u5;  
u2 WITH u6;  
u3 WITH u7;  
u4 WITH u8;
```

RESTRUCTURED ROUTINES FOR CONTINUOUS-TIME SURVIVAL ANALYSIS WITH LATENT VARIABLES

The restructuring of routines for continuous-time survival analysis with latent variables has resulted in changes to the SURVIVAL and BASEHAZARD options (see Asparouhov, 2014, Section 9). These are described below.

The SURVIVAL option is used to identify the variables that contain information about time to event and to provide information about the number and lengths of the time intervals in the baseline hazard function to be used in the analysis. The SURVIVAL option must be used in conjunction with the TIMECENSORED option. The SURVIVAL option can be specified in five ways: the default baseline hazard function, a non-parametric baseline hazard function, a semi-parametric baseline hazard function, a parametric baseline hazard function, and a constant baseline hazard function.

The SURVIVAL option is specified as follows when using the default baseline hazard function:

`SURVIVAL = t;`

where `t` is the variable that contains time-to-event information. The default is either a semi-parametric baseline hazard function with ten time intervals or a non-parametric baseline hazard function. The default is a semi-parametric baseline hazard function with ten time intervals for models where `t` is regressed on a continuous latent variable, for multilevel models, and for models that require Monte Carlo numerical integration. In this case, the lengths of the time intervals are selected internally in a non-parametric fashion. For all other models, the default is a non-parametric baseline hazard function as in Cox regression where the number and lengths of the time intervals are taken from the data and the baseline hazard function is saturated.

The SURVIVAL option is specified as follows when using a non-parametric baseline hazard function as in Cox regression:

`SURVIVAL = t (ALL);`

where `t` is the variable that contains time-to-event information and `ALL` is a keyword that specifies that the number and lengths of the time intervals are taken from the data and the baseline hazard is saturated. It is not recommended to use the keyword `ALL` when the `BASEHAZARD` option of the `ANALYSIS` command is `ON` because it results in a large number of baseline hazard parameters.

The `SURVIVAL` option is specified as follows when using a semi-parametric baseline hazard:

`SURVIVAL = t (10);`

where `t` is the variable that contains time-to-event information. The number in parentheses specifies that 10 intervals are used in the analysis where the lengths of the time intervals are selected internally in a non-parametric fashion.

The `SURVIVAL` option is specified as follows when using a parametric baseline hazard function:

`SURVIVAL = t (4*5 1*10);`

where `t` is the variable that contains time-to-event information. The numbers in parentheses specify that four time intervals of length five and one time interval of length ten are used in the analysis.

The `SURVIVAL` option is specified as follows when using a constant baseline hazard function:

`SURVIVAL = t (CONSTANT);`

where *t* is the variable that contains time-to-event information and **CONSTANT** is the keyword that specifies a constant baseline hazard function.

BASEHAZARD

The **BASEHAZARD** option is used in continuous-time survival analysis to specify whether the baseline hazard parameters are treated as model parameters or as auxiliary parameters. When the **BASEHAZARD** option is **OFF**, the parameters are treated as auxiliary parameters. When the **BASEHAZARD** option is **ON**, the parameters are treated as model parameters. In most cases, the default is **OFF**. For models where the time-to-event variable is regressed on a continuous latent variable, for multilevel models, and for models that require Monte Carlo numerical integration, the default is **ON**. Following is an example of how to request that baseline hazard parameters are treated as model parameters when this is not the default:

```
BASEHAZARD = ON;
```

With **TYPE=MIXTURE**, the **ON** and **OFF** settings have two alternatives, **EQUAL** and **UNEQUAL**. **EQUAL** is the default. With **EQUAL**, the baseline hazard parameters are held equal across classes. With **BASEHAZARD=OFF**, the baseline hazard parameters are held equal across classes as the default. To relax this equality, specify:

```
BASEHAZARD = ON (UNEQUAL);
```

or

```
BASEHAZARD = OFF (UNEQUAL);
```

In continuous-time survival modeling, there are as many baseline hazard parameters as there are time intervals plus one. When the BASEHAZARD option of the ANALYSIS command is ON, these parameters can be referred to in the MODEL command by adding to the name of the time-to-event variable the number sign (#) followed by a number. For example, for a time-to-event variable *t* with 5 time intervals, the six baseline hazard parameters are referred to as *t#1*, *t#2*, *t#3*, *t#4*, *t#5*, and *t#6*. In addition to the baseline hazard parameters, the time-to-event variable has a mean or an intercept depending on whether the model is unconditional or conditional. The mean or intercept is referred to by using a bracket statement, for example,

[*t*];

where *t* is the time-to-event variable.

The input for two Mplus User's Guide examples has changes because of changes to the SURVIVAL and BASEHAZARD options. Following are the new inputs for these examples.

EXAMPLE 6.20: CONTINUOUS-TIME SURVIVAL ANALYSIS USING THE COX REGRESSION MODEL

TITLE: this is an example of a continuous-time

```

survival analysis using the Cox regression
model
DATA:      FILE = ex6.20.dat;
VARIABLE:  NAMES = t x tc;
           SURVIVAL = t;
           TIMECENSORED = tc (0 = NOT 1 = RIGHT);
MODEL:     t ON x;

```

EXAMPLE 8.17: CONTINUOUS-TIME SURVIVAL MIXTURE ANALYSIS USING A COX REGRESSION MODEL

```

TITLE:     this is an example of a continuous-time
           survival mixture analysis using a Cox
           regression model
DATA:      FILE = ex8.17.dat;
VARIABLE:  NAMES = t u1-u5 x tc;
           CATEGORICAL = u1-u5;
           CLASSES = c (2);
           SURVIVAL = t;
           TIMECENSORED = tc (0 = NOT 1 = RIGHT);
ANALYSIS:  TYPE = MIXTURE;
MODEL:     %OVERALL%
           t ON x;
           c ON x;
           %c#1%
           [u1$1-u5$1];
           t ON x;
           %c#2%
           [u1$1-u5$1];

```

STRUCTURAL EQUATION MODELING (SEM) WITH NON-NORMAL DISTRIBUTIONS

Non-normal distributions for factors and observed variables are available using the DISTRIBUTION option of the ANALYSIS command in conjunction with TYPE=GENERAL (Asparouhov &

Muthén, 2014a). These new methods are experimental in that they have not been extensively used in practice.

The DISTRIBUTION option has four settings: NORMAL, SKEWNORMAL, TDISTRIBUTION, and SKEWT. The default is NORMAL. The DISTRIBUTION option can be used with only continuous variables although the analysis model can contain other types of variables. The DISTRIBUTION option cannot be used with models that require numerical integration.

The SKEWNORMAL and SKEWT settings have a special skew parameter for each observed and latent variable that is related to the skewness of the variable. It is specified by mentioning the name of the factor in squiggly brackets. For example, the skew parameter for a factor *f* is specified as {*f*}. Skew parameters have starting values of one. The TDISTRIBUTION and SKEWT settings have a special degree of freedom parameter that is related to the degrees of freedom in the t-distribution. The degree of freedom parameter is specified by putting *df* in squiggly brackets, for example, {*df*}. The degree of freedom parameter has a starting value of one. The SKEWNORMAL setting can capture skewness less than the absolute value of one, whereas the SKEWT setting has no such limitations.

A chi-square test of model fit is available for testing the H0 model against an unrestricted model of means, variances, covariances, skew, and degrees of freedom using the H1MODEL option of the OUTPUT command. This test is not provided by default because it can be computationally demanding. The H1MODEL has two settings: COVARIANCE and SEQUENTIAL. The default is COVARIANCE. Following is an example of how to specify the SEQUENTIAL setting:

```
H1MODEL (SEQUENTIAL);
```

The H1 model typically requires several random starts. The H1STARTS option of the ANALYSIS command is used to specify the number of random sets of starting values to generate in the initial stage and the number of optimizations to use in the final stage for the H1 model. The default is zero random sets of starting values in the initial stage and zero optimizations in the final stage.

Following is an example of how to specify the H1STARTS option:

```
H1STARTS = 100 20;
```

which specifies that 100 random sets of starting values are generated in the initial stage and 20 optimizations are carried out in the final stage.

NEW ORDER OF OPERATIONS FOR THE DEFINE COMMAND

The order of operations for the DEFINE command has changed. Previously transformations following the CLUSTER_MEAN, CENTER, and STANDARDIZE options did not use variables transformed using these options. For example, if an interaction variable was created after the CENTER option using the same variables, the interaction did not use centered variables. Now the interaction uses centered variables.

The statements in the DEFINE command are executed one observation at a time in the order specified with one exception. The CLUSTER_MEAN, CENTER, and STANDARDIZE options are executed in the order mentioned after all transformations specified before them in the DEFINE command and the DATA transformation commands are executed. All statements specified after these options are executed one observation at a time in the order specified. These transformations use the new values from

the CLUSTER_MEAN, CENTER, and STANDARDIZE options where applicable. Any variable listed in the NAMES option of the VARIABLE command or created in the DEFINE command can be transformed or used to create new variables. New variables created in the DEFINE command that will be used in an analysis must be listed on the USEVARIABLES list after the original variables. All statements specified after the CLUSTER_MEAN, CENTER and STANDARDIZE options must be a variable used in the analysis listed on the USEVARIABLES list.

DOUBLE DO LOOPS FOR THE DEFINE, MODEL CONSTRAINT, MODEL TEST, AND MODEL PRIORS COMMANDS

A double do loop is available for the DEFINE, MODEL CONSTRAINT, MODEL TEST, and MODEL PRIORS commands using the DO option. Following is an example of how to specify a double do loop in MODEL PRIORS:

```
MODEL PRIORS:  
DO ($,1,6) DO (#,1,3) p#_$_~IW(1200,100)
```

where the numbers in parentheses give the range of values the double do loop will use. The numbers replace the symbol preceding them.

The statement above is the same as saying:

```
P1_1~IW(1200,100)  
P1_2~IW(1200,100)  
P1_3~IW(1200,100)  
P1_4~IW(1200,100)  
P1_5~IW(1200,100)  
P1_6~IW(1200,100)
```

P2_1~IW(1200,100)
P2_2~IW(1200,100)
P2_3~IW(1200,100)
P2_4~IW(1200,100)
P2_5~IW(1200,100)
P2_6~IW(1200,100)
P3_1~IW(1200,100)
P3_2~IW(1200,100)
P3_3~IW(1200,100)
P3_4~IW(1200,100)
P3_5~IW(1200,100)
P3_6~IW(1200,100)

REFERENCES

Asparouhov, T. (2014). Continuous-time survival analysis in Mplus. Technical appendix. Los Angeles: Muthén & Muthén.

Asparouhov, T. & Muthén, B. (2014a). Structural equation models and mixture models with continuous non-normal skewed distributions. Mplus Web Notes: No. 19.

Asparouhov, T. & Muthén, B. (2014b). Residual associations in latent class and latent transition analysis. Forthcoming in Structural Equation Modeling.

Muthén, B. (2011). Applications of causally defined direct and indirect effects in mediation analysis using SEM in Mplus. Technical report. Los Angeles: Muthén & Muthén.

Muthén, B. & Asparouhov, T. (2014). Causal effects in mediation modeling: An introduction with applications to latent variables. Forthcoming in Structural Equation Modeling.