CHAPTER 12
EXAMPLES: MONTE CARLO SIMULATION STUDIES

Monte Carlo simulation studies are often used for methodological investigations of the performance of statistical estimators under various conditions. They can also be used to decide on the sample size needed for a study and to determine power (Muthén & Muthén, 2002). Monte Carlo studies are sometimes referred to as simulation studies.

Mplus has extensive Monte Carlo simulation facilities for both data generation and data analysis. Several types of data can be generated: simple random samples, clustered (multilevel) data, missing data, and data from populations that are observed (multiple groups) or unobserved (latent classes). Data generation models can include random effects, interactions between continuous latent variables, interactions between continuous latent variables and observed variables, and between categorical latent variables. Dependent variables can be continuous, censored, binary, ordered categorical (ordinal), unordered categorical (nominal), counts, or combinations of these variable types. In addition, two-part (semicontinuous) variables and time-to-event variables can be generated. Independent variables can be binary or continuous. All or some of the Monte Carlo generated data sets can be saved.

The analysis model can be different from the data generation model. For example, variables can be generated as categorical and analyzed as continuous or data can be generated as a three-class model and analyzed as a two-class model. In some situations, a special external Monte Carlo feature is needed to generate data by one model and analyze it by a different model. For example, variables can be generated using a clustered design and analyzed ignoring the clustering. Data generated outside of Mplus can also be analyzed using this special Monte Carlo feature.

Other special features that can be used with Monte Carlo simulation studies include saving parameter estimates from the analysis of real data to be used as population parameter and/or coverage values for data generation in a Monte Carlo simulation study. In addition, analysis results from each replication of a Monte Carlo simulation study can be

saved in an external file for further investigation. Chapter 19 discusses the options of the MONTECARLO command.

Monte Carlo data generation can include the following special features:

- Single or multiple group analysis for non-mixture models
- Missing data
- Complex survey data
- Latent variable interactions and non-linear factor analysis using maximum likelihood
- Random slopes
- Individually-varying times of observations
- Linear and non-linear parameter constraints
- Indirect effects including specific paths
- Maximum likelihood estimation for all outcome types
- Wald chi-square test of parameter equalities
- Analysis with between-level categorical latent variables

Multiple group data generation is specified by using the NGROUPS option of the MONTECARLO command and the MODEL POPULATION-label command. Missing data generation is specified by using the PATMISS and PATPROBS options of the MONTECARLO command or the MISSING option of the MONTECARLO command in conjunction with the MODEL MISSING command. Complex survey data are generated by using the TYPE=TWOLEVEL option of the ANALYSIS command in conjunction with the NCSIZES and CSIZES options of the MONTECARLO command. Latent variable interactions are generated by using the | symbol of the MODEL POPULATION command in conjunction with the XWITH option of the MODEL POPULATION command. Random slopes are generated by using the | symbol of the MODEL POPULATION command in conjunction with the ON option of the MODEL POPULATION command. Individually-varying times of observations are generated by using the | symbol of the MODEL POPULATION command in conjunction with the AT option of the MODEL POPULATION command and the TSCORES option of the MONTECARLO command. Linear and non-linear parameter constraints are specified by using the MODEL CONSTRAINT command. Indirect effects are specified by using the MODEL INDIRECT command. Maximum likelihood estimation is specified by using the ESTIMATOR option of the ANALYSIS command. The MODEL TEST command is used to test linear restrictions on the parameters in the MODEL and

MODEL CONSTRAINT commands using the Wald chi-square test. Between-level categorical latent variables are generated using the GENCLASSES option and specified using the CLASSES and BETWEEN options.

Besides the examples in this chapter, Monte Carlo versions of most of the examples in the previous example chapters are included on the CD that contains the Mplus program and at www.statmodel.com. Following is the set of Monte Carlo examples included in this chapter:

- 12.1: Monte Carlo simulation study for a CFA with covariates (MIMIC) with continuous factor indicators and patterns of missing data
- 12.2: Monte Carlo simulation study for a linear growth model for a continuous outcome with missing data where attrition is predicted by time-invariant covariates (MAR)
- 12.3: Monte Carlo simulation study for a growth mixture model with two classes and a misspecified model
- 12.4: Monte Carlo simulation study for a two-level growth model for a continuous outcome (three-level analysis)
- 12.5: Monte Carlo simulation study for an exploratory factor analysis with continuous factor indicators
- 12.6 Step 1: Monte Carlo simulation study where clustered data for a two-level growth model for a continuous outcome (three-level analysis) are generated, analyzed, and saved
- 12.6 Step 2: External Monte Carlo analysis of clustered data generated for a two-level growth model for a continuous outcome using TYPE=COMPLEX for a single-level growth model
- 12.7 Step 1: Real data analysis of a CFA with covariates (MIMIC) for continuous factor indicators where the parameter estimates are saved for use in a Monte Carlo simulation study
- 12.7 Step 2: Monte Carlo simulation study where parameter estimates saved from a real data analysis are used for population parameter values for data generation and coverage
- 12.8: Monte Carlo simulation study for discrete-time survival analysis*
- 12.9: Monte Carlo simulation study for a two-part (semicontinuous) growth model for a continuous outcome*
- 12.10: Monte Carlo simulation study for a two-level continuous-time survival analysis using Cox regression with a random intercept and a frailty*

- 12.11: Monte Carlo simulation study for a two-level mediation model with random slopes
- 12.12 Monte Carlo simulation study for a multiple group EFA with continuous factor indicators with measurement invariance of intercepts and factor loadings

* Example uses numerical integration in the estimation of the model. This can be computationally demanding depending on the size of the problem.

MONTE CARLO DATA GENERATION

Data are generated according to the following steps. First, multivariate normal data are generated for the independent variables in the model. Second, the independent variables are categorized if requested. The third step varies depending on the dependent variable type and the model used. Data for continuous dependent variables are generated according to a distribution that is multivariate normal conditional on the independent variables. For categorical dependent variables under the probit model using weighted least squares estimation, data for continuous dependent variables are generated according to a distribution that is multivariate normal conditional on the independent variables. These dependent variables are then categorized using the thresholds provided in the MODEL POPULATION command or the POPULATION option of the MONTECARLO command. For categorical dependent variables under the probit model using maximum likelihood estimation, the dependent variables are generated according to the probit model using the values of the thresholds and slopes from the MODEL POPULATION command or the POPULATION option of the MONTECARLO command. For categorical dependent variables under the logistic model using maximum likelihood estimation, the dependent variables are generated according to the logistic model using the values of the thresholds and slopes from the MODEL POPULATION command or the POPULATION option of the MONTECARLO command. For censored dependent variables, the dependent variables are generated according to the censored normal model using the values of the intercepts and slopes from the MODEL POPULATION command or the POPULATION option of the MONTECARLO command. For unordered categorical (nominal) dependent variables, the dependent variables are generated according to the multinomial logistic model using the values of the intercepts and slopes from the MODEL

POPULATION command or the POPULATION option of the MONTECARLO command. For count dependent variables, the dependent variables are generated according to the log rate model using the values of the intercepts and slopes from the MODEL POPULATION command or the POPULATION option of the MONTECARLO command. For time-to-event variables in continuous-time survival analysis, the dependent variables are generated according to the loglinear model using the values of the intercepts and slopes from the MODEL POPULATION command or the POPULATION option of the MONTECARLO command.

To save the generated data for subsequent analysis without analyzing them, use the TYPE=BASIC option of the ANALYSIS command in conjunction with the REPSAVE and SAVE options of the MONTECARLO command.

MONTE CARLO DATA ANALYSIS

There are two ways to carry out a Monte Carlo simulation study in Mplus: an internal Monte Carlo simulation study or an external Monte Carlo simulation study. In an internal Monte Carlo simulation study, data are generated and analyzed in one step using the MONTECARLO command. In an external Monte Carlo simulation study, multiple data sets are generated in a first step using either Mplus or another computer program. These data are analyzed and the results summarized in a second step using regular Mplus analysis facilities in conjunction with the TYPE=MONTECARLO option of the DATA command.

Internal Monte Carlo can be used whenever the analysis type and scales of the dependent variables remain the same for both data generation and analysis. Internal Monte Carlo can also be used with TYPE=GENERAL when dependent variables are generated as categorical and analyzed as continuous. Internal Monte Carlo can also be used when data are generated and analyzed for a different number of latent classes. In all other cases, data from all replications can be saved and subsequently analyzed using external Monte Carlo.

MONTE CARLO OUTPUT

The default output for the MONTECARLO command includes a listing of the input setup, a summary of the analysis specifications, sample statistics from the first replication, the analysis results summarized over replications, and TECH1 which shows the free parameters in the model and the starting values. Following is an example of the output for tests of model fit for the chi-square test statistic. The same format is used with other fit statistics.

```
Chi-Square Test of Model Fit

        Degrees of freedom                          5

        Mean                                    5.253
        Std Dev                                 3.325
        Number of successful computations         500

           Proportions                    Percentiles
        Expected    Observed          Expected        Observed
          0.990       0.988             0.554           0.372
          0.980       0.976             0.752           0.727
          0.950       0.958             1.145           1.193
          0.900       0.894             1.610           1.539
          0.800       0.804             2.343           2.367
          0.700       0.710             3.000           3.090
          0.500       0.532             4.351           4.555
          0.300       0.330             6.064           6.480
          0.200       0.242             7.289           7.870
          0.100       0.136             9.236           9.950
          0.050       0.062            11.070          11.576
          0.020       0.022            13.388          13.394
          0.010       0.014            15.086          15.146
```

The mean and standard deviation of the chi-square test statistic over the replications of the Monte Carlo analysis are given. The column labeled Proportions Expected (column 1) should be understood in conjunction with the column labeled Percentiles Expected (column 3). Each value in column 1 gives the probability of observing a chi-square value greater than the corresponding value in column 3. The column 3 percentile values are determined from a chi-square distribution with the degrees of freedom given by the model, in this case 5. In this output, the column 1 value of 0.05 gives the probability that the chi-square value exceeds the column 3 percentile value (the critical value of the chi-square distribution) of 11.070. Columns 2 and 4 give the corresponding values observed in the Monte Carlo replications. Column 2 gives the proportion of replications for which the critical value is exceeded, which

in this example is 0.062, close to the expected value 0.05 which indicates that the chi-square distribution is well approximated in this case. The column 4 value of 11.576 is the chi-square value at this percentile from the Monte Carlo analysis that has 5% of the values in the replications above it. The fact that it deviates little from the theoretical value of 11.070 is again an indication that the chi-square distribution is well approximated in this case. For the other fit statistics, the normal distribution is used to obtain the critical values of the test statistic.

The summary of the analysis results includes the population value for each parameter, the average of the parameter estimates across replications, the standard deviation of the parameter estimates across replications, the average of the estimated standard errors across replications, the mean square error for each parameter (M.S.E.), 95 percent coverage, and the proportion of replications for which the null hypothesis that a parameter is equal to zero is rejected at the .05 level.

```
MODEL RESULTS

                            ESTIMATES              S. E.    M. S. E.   95%   % Sig
            Population   Average   Std. Dev.   Average               Cover  Coeff
 I        |
  Y1             1.000    1.0000    0.0000      0.0000      0.0000  1.000  0.000
  Y2             1.000    1.0000    0.0000      0.0000      0.0000  1.000  0.000
  Y3             1.000    1.0000    0.0000      0.0000      0.0000  1.000  0.000
  Y4             1.000    1.0000    0.0000      0.0000      0.0000  1.000  0.000

 S        |
  Y1             0.000    0.0000    0.0000      0.0000      0.0000  1.000  0.000
  Y2             1.000    1.0000    0.0000      0.0000      0.0000  1.000  0.000
  Y3             2.000    2.0000    0.0000      0.0000      0.0000  1.000  0.000
  Y4             3.000    3.0000    0.0000      0.0000      0.0000  1.000  0.000

 I        WITH
  S              0.000    0.0006    0.0301      0.0306      0.0009  0.958  0.042

 Means
  I              0.000   -0.0006    0.0473      0.0460      0.0022  0.950  0.050
  S              0.200    0.2015    0.0278      0.0274      0.0008  0.946  1.000

 Variances
  I              0.500    0.4969    0.0704      0.0685      0.0050  0.936  1.000
  S              0.200    0.1997    0.0250      0.0237      0.0006  0.930  1.000

 Residual Variances
  Y1             0.500    0.5016    0.0683      0.0657      0.0047  0.934  1.000
  Y2             0.500    0.5018    0.0460      0.0451      0.0021  0.958  1.000
  Y3             0.500    0.5025    0.0515      0.0532      0.0027  0.956  1.000
  Y4             0.500    0.4991    0.0932      0.0918      0.0087  0.946  1.000
```

The column labeled Population gives the population parameter values that are given in the MODEL command, the MODEL COVERAGE command, or using the COVERAGE option of the MONTECARLO command. The column labeled Average gives the average of the parameter estimates across the replications of the Monte Carlo simulation study. These two values are used to evaluate parameter bias. To determine the percentage of parameter bias, subtract the population parameter value from the average parameter value, divide this number by the population parameter value, and multiply by 100. The parameter bias for the variance of i would be

100 (.4969 - .5000) / .5000 = -0.62.

This results in a bias of -0.62 percent.

The column labeled Std. Dev. gives the standard deviation of the parameter estimates across the replications of the Monte Carlo simulation study. When the number of replications is large, this is considered to be the population standard error. The column labeled S.E. Average gives the average of the estimated standard errors across replications of the Monte Carlo simulation study. To determine standard error bias, subtract the population standard error value from the average standard error value, divide this number by the population standard error value, and multiply by 100.

The column labeled M.S.E. gives the mean square error for each parameter. M.S.E. is equal to the variance of the estimates across the replications plus the square of the bias. For example, the M.S.E. for the variance of i is equal to 0.0704 squared plus (0.4969 - 0.5) squared which is equal to 0.00497 or 0.0050. The column labeled 95% Cover gives the proportion of replications for which the 95% confidence interval contains the population parameter value. This gives the coverage which indicates how well the parameters and their standard errors are estimated. In this output, all coverage values are close to the correct value of 0.95.

The column labeled % Sig Coeff gives the proportion of replications for which the null hypothesis that a parameter is equal to zero is rejected at the .05 level (two-tailed test with a critical value of 1.96). The statistical test is the ratio of the parameter estimate to its standard error, an approximately normally distributed quantity (z-score) in large samples.

For parameters with population values different from zero, this value is an estimate of power with respect to a single parameter, that is, the probability of rejecting the null hypothesis when it is false. For parameters with population values equal to zero, this value is an estimate of Type I error, that is, the probability of rejecting the null hypothesis when it is true. In this output, the power to reject that the slope growth factor mean is zero is estimated as 1.000, that is, exceeding the standard of 0.8 power.

MONTE CARLO EXAMPLES

Following is the set of Monte Carlo simulation study examples. Besides the examples in this chapter, Monte Carlo versions of most of the examples in the previous example chapters are included on the CD that contains the Mplus program and at www.statmodel.com.

EXAMPLE 12.1: MONTE CARLO SIMULATION STUDY FOR A CFA WITH COVARIATES (MIMIC) WITH CONTINUOUS FACTOR INDICATORS AND PATTERNS OF MISSING DATA

```
TITLE:      this is an example of a Monte Carlo
            simulation study for a CFA with covariates
            (MIMIC) with continuous factor indicators
            and patterns of missing data
MONTECARLO:
            NAMES ARE y1-y4 x1 x2;
            NOBSERVATIONS = 500;
            NREPS = 500;
            SEED = 4533;
            CUTPOINTS = x2(1);
            PATMISS = y1(.1) y2(.2) y3(.3) y4(1) |
                      y1(1) y2(.1) y3(.2) y4(.3);
            PATPROBS = .4 | .6;
MODEL POPULATION:
            [x1-x2@0];
            x1-x2@1;
            f BY y1@1 y2-y4*1;
            f*.5;
            y1-y4*.5;
            f ON x1*1 x2*.3;
MODEL:      f BY y1@1 y2-y4*1;
            f*.5;
            y1-y4*.5;
            f ON x1*1 x2*.3;
OUTPUT:     TECH9;
```

In this example, data are generated and analyzed according to the CFA with covariates (MIMIC) model described in Example 5.8. Two factors are regressed on two covariates and data are generated with patterns of missing data.

```
TITLE:      this  is  an  example  of  a  Monte  Carlo
            simulation study for a CFA with covariates
            (MIMIC) with continuous factor indicators
            and patterns of missing data
```

The TITLE command is used to provide a title for the output. The title is printed in the output just before the Summary of Analysis.

```
MONTECARLO:
        NAMES ARE y1-y4 x1 x2;
        NOBSERVATIONS = 500;
        NREPS = 500;
        SEED = 4533;
        CUTPOINTS = x2(1);
        PATMISS = y1(.1) y2(.2) y3(.3) y4(1) |
                  y1(1) y2(.1) y3(.2) y4(.3);
        PATPROBS = .4 | .6;
```

The MONTECARLO command is used to describe the details of a Monte Carlo simulation study. The NAMES option is used to assign names to the variables in the generated data sets. The data sets in this example each have six variables: y1, y2, y3, y4, x1, and x2. Note that a hyphen can be used as a convenience feature in order to generate a list of names. The NOBSERVATIONS option is used to specify the sample size to be used for data generation and for analysis. In this example, the sample size is 500. The NREPS option is used to specify the number of replications, that is, the number of samples to draw from a specified population. In this example, 500 samples will be drawn. The SEED option is used to specify the seed to be used for the random draws. The seed 4533 is used here. The default seed value is zero.

The GENERATE option is used to specify the scale of the dependent variables for data generation. In this example, the dependent variables are continuous which is the default for the GENERATE option. Therefore, the GENERATE option is not necessary and is not used here. The CUTPOINTS option is used to create binary variables from the multivariate normal independent variables generated by the program. In this example, the variable x2 is cut at the value of one which is one standard deviation above the mean because the mean and variance used for data generation are zero and one. This implies that after the cut x2 is a 0/1 binary variable where 16 percent of the population have the value of 1. The mean and variance of x2 for data generation are specified in the MODEL POPULATION command.

The PATMISS and PATPROBS options are used together to describe the patterns of missing data to be used in data generation. The PATMISS option is used to specify the missing data patterns and the proportion missing for each variable. The patterns are separated using the | symbol. The PATPROBS option is used to specify the proportion of individuals for each missing data pattern. In this example, there are two missing value patterns. In the first pattern, y1 has 10 percent

missing, y2 has 20 percent missing, y3 has 30 percent missing, and y4 has 100 percent missing. In the second pattern, y1 has 100 percent missing, y2 has 10 percent missing, y3 has 20 percent missing, and y4 has 30 percent missing. As specified in the PATPROBS option, 40 percent of the individuals in the generated data have missing data pattern 1 and 60 percent have missing data pattern 2. This may correspond to a situation of planned missingness where a measurement instrument is administered in two different versions given to randomly chosen parts of the population. In this example, some individuals answer items y1, y2, and y3, while others answer y2, y3, and y4.

```
MODEL POPULATION:
          [x1-x2@0];
          x1-x2@1;
          f BY y1@1 y2-y4*1;
          f*.5;
          y1-y4*.5;
          f ON x1*1 x2*.3;
```

The MODEL POPULATION command is used to provide the population parameter values to be used in data generation. Each parameter in the model must be specified followed by the @ symbol or the asterisk (*) and the population parameter value. Any model parameter not given a population parameter value will be assigned the value of zero as the population parameter value. The first two lines in the MODEL POPULATION command refer to the means and variances of the independent variables x1 and x2. The covariances between the independent variables can also be specified. Variances of the independent variables in the model must be specified. Means and covariances of the independent variables do not need to be specified if their values are zero.

```
MODEL:     f BY y1@1 y2-y4*1;
           f*.5;
           y1-y4*.5;
           f ON x1*1 x2*.3;
```

The MODEL command is used to describe the analysis model as in regular analyses. In Monte Carlo simulation studies, the MODEL command is also used to provide values for each parameter that are used as population parameter values for computing coverage and mean square error and starting values in the estimation of the model. They are printed in the first column of the output labeled Population. Population

parameter values for the analysis model can also be provided using the MODEL COVERAGE command or the COVERAGE option of the MONTECARLO command. Alternate starting values can be provided using the STARTING option of the MONTECARLO command. Note that the population parameter values for coverage given in the analysis model are different from the population parameter values used for data generation if the analysis model is misspecified.

```
OUTPUT:    TECH9;
```

The OUTPUT command is used to request additional output not included as the default. The TECH9 option is used to request error messages related to convergence for each replication of the Monte Carlo simulation study.

EXAMPLE 12.2: MONTE CARLO SIMULATION STUDY FOR A LINEAR GROWTH MODEL FOR A CONTINUOUS OUTCOME WITH MISSING DATA WHERE ATTRITION IS PREDICTED BY TIME-INVARIANT COVARIATES (MAR)

```
TITLE:     this is an example of a Monte Carlo
           simulation study for a linear growth model
           for a continuous outcome with missing data
           where attrition is predicted by time-
           invariant covariates (MAR)
MONTECARLO:
           NAMES ARE y1-y4 x1 x2;
           NOBSERVATIONS = 500;
           NREPS = 500;
           SEED = 4533;
           CUTPOINTS = x2(1);
           MISSING = y1-y4;
MODEL POPULATION:
           x1-x2@1;
           [x1-x2@0];
           i s | y1@0 y2@1 y3@2 y4@3;
           [i*1 s*2];
           i*1; s*.2; i WITH s*.1;
           y1-y4*.5;
           i ON x1*1 x2*.5;
           s ON x1*.4 x2*.25;
```

```
MODEL MISSING:
            [y1-y4@-1];
            y1 ON x1*.4 x2*.2;
            y2 ON x1*.8 x2*.4;
            y3 ON x1*1.6 x2*.8;
            y4 ON x1*3.2 x2*1.6;
MODEL:      i s | y1@0 y2@1 y3@2 y4@3;
            [i*1 s*2];
            i*1; s*.2; i WITH s*.1;
            y1-y4*.5;
            i ON x1*1 x2*.5;
            s ON x1*.4 x2*.25;
OUTPUT:     TECH9;
```

In this example, missing data are generated to illustrate both random missingness and attrition predicted by time-invariant covariates (MAR). This Monte Carlo simulation study can be used to estimate the power to detect that the binary covariate x2 has a significant effect on the growth slope factor s. The binary covariate x2 may correspond to a treatment variable or a gender variable.

The MISSING option in the MONTECARLO command is used to identify the dependent variables in the data generation model for which missing data will be generated. The MODEL MISSING command is used to provide information about the population parameter values for the missing data model to be used in the generation of data. The MODEL MISSING command specifies a logistic regression model for a set of binary dependent variables that represent not missing (scored as 0) and missing (scored as 1) for the dependent variables in the data generation model. The first statement in the MODEL MISSING command defines the intercepts in the logistic regressions for each of the binary dependent variables. If the covariates predicting missingness all have values of zero, the logistic regression intercept value of -1 corresponds to a probability of 0.27 of having missing data on the dependent variables. This would reflect missing completely at random. The four ON statements specify the logistic regression of the four binary dependent variables on the two covariates x1 and x2 to reflect attrition predicted by the covariates. Because the values of the logistic regression slopes increase over time as seen in the increase of the slopes from y1 to y4, attrition also increases over time and becomes more selective over time. An explanation of the other commands can be found in Example 12.1.

EXAMPLE 12.3: MONTE CARLO SIMULATION STUDY FOR A GROWTH MIXTURE MODEL WITH TWO CLASSES AND A MISSPECIFIED MODEL

```
TITLE:      this is an example of a Monte Carlo
            simulation study for a growth mixture
            model with two classes and a misspecified
            model
MONTECARLO:
            NAMES ARE u y1-y4 x;
            NOBSERVATIONS = 500;
            NREPS = 10;
            SEED = 53487;
            GENERATE = u (1);
            CATEGORICAL = u;
            GENCLASSES = c (2);
            CLASSES = c (1);
MODEL POPULATION:
            %OVERALL%
            [x@0];
            x@1;
            i s | y1@0 y2@1 y3@2 y4@3;
            i*.25 s*.04;
            i WITH s*0;
            y1*.4 y2*.35 y3*.3 y4*.25;
            i ON x*.5;
            s ON x*.1;
            c#1 ON x*.2;
            [c#1*0];
            %c#1%
            [u$1*1 i*3 s*.5];
            %c#2%
            [u$1*-1 i*1 s*0];
```

```
ANALYSIS:  TYPE = MIXTURE;
MODEL:
           %OVERALL%
           i s | y1@0 y2@1 y3@2 y4@3;
           i*.25 s*.04;
           i WITH s*0;
           y1*.4 y2*.35 y3*.3 y4*.25;
           i ON x*.5;
           s ON x*.1;
!          c#1 ON x*.2;
!          [c#1*0];
           u ON x;
           %c#1%
           [u$1*1 i*3 s*.5];
!          %c#2%
!          [u$1*-1 i*1 s*0];
OUTPUT:    TECH9;
```

In this example, data are generated according the two class model described in Example 8.1 and analyzed as a one class model. This results in a misspecified model. Differences between the parameter values that generated the data and the estimated parameters can be studied to determine the extent of the distortion.

The GENERATE option is used to specify the scale of the dependent variables for data generation. In this example, the dependent variable u is binary because it has one threshold. For binary variables, this is specified by placing the number one in parenthesis following the variable name. The CATEGORICAL option is used to specify which dependent variables are treated as binary or ordered categorical (ordinal) variables in the model and its estimation. In the example above, the variable u is generated and analyzed as a binary variable. The GENCLASSES option is used to assign names to the categorical latent variables in the data generation model and to specify the number of latent classes to be used for data generation. In the example above, there is one categorical latent variable c that has two latent classes for data generation. The CLASSES option is used to assign names to the categorical latent variables in the analysis model and to specify the number of latent classes to be used for analysis. In the example above, there is one categorical latent variable c that has one latent class for analysis. The ANALYSIS command is used to describe the technical details of the analysis. The TYPE option is used to describe the type of analysis that is to be performed. By selecting MIXTURE, a mixture model will be estimated.

The commented out lines in the MODEL command show how the MODEL command is changed from a two class model to a one class model. An explanation of the other commands can be found in Examples 12.1 and 8.1.

EXAMPLE 12.4: MONTE CARLO SIMULATION STUDY FOR A TWO-LEVEL GROWTH MODEL FOR A CONTINUOUS OUTCOME (THREE-LEVEL ANALYSIS)

```
TITLE:     this is an example of a Monte Carlo
           simulation study for a two-level growth
           model for a continuous outcome (three-
           level analysis)
MONTECARLO:
           NAMES ARE y1-y4 x w;
           NOBSERVATIONS = 1000;
           NREPS = 500;
           SEED = 58459;
           CUTPOINTS = x (1) w (0);
           MISSING = y1-y4;
           NCSIZES = 3;
           CSIZES = 40 (5) 50 (10) 20 (15);
           WITHIN = x;
           BETWEEN = w;
MODEL POPULATION:
           %WITHIN%
           x@1;
           iw sw | y1@0 y2@1 y3@2 y4@3;
           y1-y4*.5;
           iw ON x*1;
           sw ON x*.25;
           iw*1; sw*.2;
           %BETWEEN%
           w@1;
           ib sb | y1@0 y2@1 y3@2 y4@3;
           y1-y4@0;
           ib ON w*.5;
           sb ON w*.25;
           [ib*1 sb*.5];
           ib*.2; sb*.1;
```

```
MODEL MISSING:
            [y1-y4@-1];
            y1 ON x*.4;
            y2 ON x*.8;
            y3 ON x*1.6;
            y4 ON x*3.2;
ANALYSIS:  TYPE IS TWOLEVEL;
MODEL:
            %WITHIN%
            iw sw | y1@0 y2@1 y3@2 y4@3;
            y1-y4*.5;
            iw ON x*1;
            sw ON x*.25;
            iw*1; sw*.2;
            %BETWEEN%
            ib sb | y1@0 y2@1 y3@2 y4@3;
            y1-y4@0;
            ib ON w*.5;
            sb ON w*.25;
            [ib*1 sb*.5];
            ib*.2; sb*.1;
OUTPUT:    TECH9 NOCHISQUARE;
```

In this example, data for the two-level growth model for a continuous outcome (three-level analysis) described in Example 9.12 are generated and analyzed. This Monte Carlo simulation study can be used to estimate the power to detect that the binary cluster-level covariate w has a significant effect on the growth slope factor sb.

The NCSIZES option is used to specify the number of unique cluster sizes to be used in data generation. In the example above, there are three unique cluster sizes. The CSIZES option is used to specify the number of clusters and the sizes of the clusters to be used in data generation. The CSIZES option specifies that 40 clusters of size 5, 50 clusters of size 10, and 20 clusters of size 15 will be generated. The WITHIN option is used to identify the variables in the data set that are measured on the individual level and modeled only on the within level. They are specified to have no variance in the between part of the model. The variable x is an individual-level variable. The BETWEEN option is used to identify the variables in the data set that are measured on the cluster level and modeled only on the between level. The variable w is a cluster-level variable. Variables not mentioned on the WITHIN or the BETWEEN statements are measured on the individual level and can be modeled on both the within and between levels. The NOCHISQUARE option of the OUTPUT command is used to request that the chi-square

fit statistic not be computed. This reduces computational time. An explanation of the other commands can be found in Examples 12.1 and 12.2 and Example 9.12.

EXAMPLE 12.5: MONTE CARLO SIMULATION STUDY FOR AN EXPLORATORY FACTOR ANALYSIS WITH CONTINUOUS FACTOR INDICATORS

```
TITLE:      this is an example of a Monte Carlo
            simulation study for an exploratory factor
            analysis with continuous factor indicators
MONTECARLO:
            NAMES ARE y1-y10;
            NOBSERVATIONS = 500;
            NREPS = 500;
MODEL POPULATION:
            f1 BY y1-y7*.5;
            f2 BY y4-y5*.25 y6-y10*.8;
            f1-f2@1;
            f1 WITH f2*.5;
            y1-y10*.36;
MODEL:      f1 BY y1-y7*.5 y8-y10*0 (*1);
            f2 BY y1-y3*.0 y4-y5*.25 y6-y10*.8 (*1);
            f1 WITH f2*.5;
            y1-y10*.36;
OUTPUT:     TECH9;
```

In this example, data are generated according to a two-factor CFA model with continuous outcomes and analyzed as an exploratory factor analysis using exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009a).

In the MODEL command, the BY statements specify that the factors f1 and f2 are measured by the continuous factor indicators y1 through y10. The label 1 following an asterisk (*) in parentheses following the BY statements is used to indicate that f1 and f2 are a set of EFA factors. When no rotation is specified using the ROTATION option of the ANALYSIS command, the default oblique GEOMIN rotation is used to obtain factor loadings and factor correlations. The intercepts and residual variances of the factor indicators are estimated and the residuals are not correlated as the default. The variances of the factors are fixed at

one as the default. The factors are correlated under the default oblique GEOMIN rotation.

The default estimator for this type of analysis is maximum likelihood. The ESTIMATOR option of the ANALYSIS command can be used to select a different estimator. An explanation of the other commands can be found in Examples 12.1 and 12.2.

EXAMPLE 12.6 STEP 1: MONTE CARLO SIMULATION STUDY WHERE CLUSTERED DATA FOR A TWO-LEVEL GROWTH MODEL FOR A CONTINUOUS OUTCOME (THREE-LEVEL ANALYSIS) ARE GENERATED, ANALYZED, AND SAVED

```
TITLE:      this is an example of a Monte Carlo
            simulation study where clustered data for
            a two-level growth model for a continuous
            outcome (three-level) analysis are
            generated and analyzed
MONTECARLO:
            NAMES ARE y1-y4 x w;
            NOBSERVATIONS = 1000;
            NREPS = 100;
            SEED = 58459;
            CUTPOINTS = x(1) w(0);
            MISSING = y1-y4;
            NCSIZES = 3;
            CSIZES = 40 (5) 50 (10) 20 (15);
            WITHIN = x;  BETWEEN = w;
            REPSAVE = ALL;
            SAVE = ex12.6rep*.dat;
MODEL POPULATION:
            %WITHIN%
            x@1;
            iw sw | y1@0 y2@1 y3@2 y4@3;
            y1-y4*.5;
            iw ON x*1;
            sw ON x*.25;
            iw*1; sw*.2;
            %BETWEEN%
            w@1;
            ib sb | y1@0 y2@1 y3@2 y4@3;
            y1-y4@0;
            ib ON w*.5;
```

```
             sb ON w*.25;
             [ib*1 sb*.5];
             ib*.2; sb*.1;
MODEL MISSING:
             [y1-y4@-1];
             y1 ON x*.4;
             y2 ON x*.8;
             y3 ON x*1.6;
             y4 ON x*3.2;
ANALYSIS:   TYPE = TWOLEVEL;
MODEL:
             %WITHIN%
             iw sw | y1@0 y2@1 y3@2 y4@3;
             y1-y4*.5;
             iw ON x*1;
             sw ON x*.25;
             iw*1; sw*.2;
             %BETWEEN%
             ib sb | y1@0 y2@1 y3@2 y4@3;
             y1-y4@0;
             ib ON w*.5;
             sb ON w*.25;
             [ib*1 sb*.5];
             ib*.2; sb*.1;
OUTPUT:     TECH8 TECH9;
```

In this example, clustered data are generated and analyzed for the two-level growth model for a continuous outcome (three-level) analysis described in Example 9.12. The data are saved for a subsequent external Monte Carlo simulation study. The REPSAVE and SAVE options of the MONTECARLO command are used to save some or all of the data sets generated in a Monte Carlo simulation study. The REPSAVE option specifies the numbers of the replications for which the data will be saved. In the example above, the keyword ALL specifies that all of the data sets will be saved. The SAVE option is used to name the files to which the data sets will be written. The asterisk (*) is replaced by the replication number. For example, data from the first replication will be saved in the file named ex12.6rep1.dat. A file is also produced where the asterisk (*) is replaced by the word list. The file, in this case ex12.6replist.dat, contains the names of the generated data sets. The ANALYSIS command is used to describe the technical details of the analysis. By selecting TYPE=TWOLEVEL, a multilevel model is estimated. An explanation of the other commands can be found in Examples 12.1, 12.2, 12.4 and Example 9.12.

EXAMPLE 12.6 STEP 2: EXTERNAL MONTE CARLO ANALYSIS OF CLUSTERED DATA GENERATED FOR A TWO-LEVEL GROWTH MODEL FOR A CONTINUOUS OUTCOME USING TYPE=COMPLEX FOR A SINGLE-LEVEL GROWTH MODEL

```
TITLE:      this is an example of an external Monte
            Carlo analysis of clustered data generated
            for a two-level growth model for a
            continuous outcome using TYPE=COMPLEX for
            a single-level growth model
DATA:       FILE = ex12.6replist.dat;
            TYPE = MONTECARLO;
VARIABLE:   NAMES = y1-y4 x w clus;
            USEVARIABLES = y1-w;
            MISSING = ALL (999);
            CLUSTER = clus;
ANALYSIS:   TYPE = COMPLEX;
MODEL:      i s | y1@0 y2@1 y3@2 y4@3;
            y1-y4*.5;
            i ON x*1 w*.5;
            s ON x*.25 w*.25;
            i*1.2; s*.3;
            [i*1 s*.5];
OUTPUT:     TECH9;
```

In this example, an external Monte Carlo simulation study of clustered data generated for a two-level growth model for a continuous outcome is carried out using TYPE=COMPLEX for a single-level growth model. The DATA command is used to provide information about the data sets to be analyzed. The MONTECARLO setting of the TYPE option is used when the data sets being analyzed have been generated and saved using either the REPSAVE option of the MONTECARLO command or by another computer program. The file named using the FILE option of the DATA command contains a list of the names of the data sets to be analyzed and summarized as in a Monte Carlo simulation study. This file is created when the SAVE and REPSAVE options of the MONTECARLO command are used to save Monte Carlo generated data sets. The CLUSTER option of the VARIABLE command is used when data have been collected under a complex survey data design to identify the variable that contains cluster information. In the example above, the variable clus contains cluster information.   By selecting

TYPE=COMPLEX, an analysis is carried out that takes non-independence of observations into account.

In external Monte Carlo simulation studies, the MODEL command is also used to provide values for each parameter. These are used as the population parameter values for the analysis model and are printed in the first column of the output labeled Population. They are used for computing coverage and as starting values in the estimation of the model.

EXAMPLE 12.7 STEP 1: REAL DATA ANALYSIS OF A CFA WITH COVARIATES (MIMIC) FOR CONTINUOUS FACTOR INDICATORS WHERE THE PARAMETER ESTIMATES ARE SAVED FOR USE IN A MONTE CARLO SIMULATION STUDY

```
TITLE:     this is an example of a real data analysis
           of a CFA with covariates (MIMIC) for
           continuous factor indicators where the
           parameter estimates are saved for use in a
           Monte Carlo simulation study
DATA:      FILE = ex12.7real.dat;
VARIABLE:  NAMES = y1-y10 x1 x2;
MODEL:     f1 BY y1@1 y2-y5*1;
           f2 BY y6@1 y7-y10*1;
           f1-f2*.5;
           f1 WITH f2*.25;
           y1-y5*.5;
           [y1-y5*1];
           y6-y10*.75;
           [y6-y10*2];
           f1 ON x1*.3 x2*.5;
           f2 ON x1*.5 x2*.3;
OUTPUT:    TECH1;
SAVEDATA:  ESTIMATES = ex12.7estimates.dat;
```

In this example, parameter estimates from a real data analysis of a CFA with covariates (MIMIC) for continuous factor indicators are saved for use as population parameter values for use in data generation and coverage in a subsequent internal Monte Carlo simulation study. The ESTIMATES option of the SAVEDATA command is used to specify the name of the file in which the parameter estimates of the analysis will be saved.

EXAMPLE 12.7 STEP 2: MONTE CARLO SIMULATION STUDY WHERE PARAMETER ESTIMATES SAVED FROM A REAL DATA ANALYSIS ARE USED FOR POPULATION PARAMETER VALUES FOR DATA GENERATION AND COVERAGE

```
TITLE:     this is an example of a Monte Carlo
           simulation study where parameter estimates
           saved from a real data analysis are used
           for population parameter values for data
           generation and coverage
MONTECARLO:
           NAMES ARE y1-y10 x1 x2;
           NOBSERVATIONS = 500;
           NREPS = 500;
           SEED = 45335;
           POPULATION = ex12.7estimates.dat;
           COVERAGE = ex12.7estimates.dat;
MODEL POPULATION:
           f1 BY y1-y5;
           f2 BY y6-y10;
           f1 ON x1 x2;
           f2 ON x1 x2;
MODEL:     f1 BY y1-y5;
           f2 BY y6-y10;
           f1 ON x1 x2;
           f2 ON x1 x2;
OUTPUT:    TECH9;
```

In this example, parameter estimates saved from a real data analysis are used for population parameter values for data generation and coverage using the POPULATION and COVERAGE options of the MONTECARLO command. The POPULATION option is used to name the data set that contains the population parameter values to be used in data generation. The COVERAGE option is used to name the data set that contains the parameter values to be used for computing coverage and are printed in the first column of the output labeled Population. An explanation of the other commands can be found in Example 12.1.

EXAMPLE 12.8: MONTE CARLO SIMULATION STUDY FOR DISCRETE-TIME SURVIVAL ANALYSIS

```
TITLE:      this is an example of a Monte Carlo
            simulation study for discrete-time
            survival analysis
MONTECARLO:
            NAMES = u1-u4 x;
            NOBSERVATIONS = 1000;
            NREPS = 100;
            GENERATE = u1-u4(1);
            MISSING = u2-u4;
            CATEGORICAL = u1-u4;
MODEL POPULATION:
            [x@0]; x@1;
            [u1$1*2 u2$1*1.5 u3$1*1 u4$1*1];
            f BY u1-u4@1;
            f ON x*.5;
            f@0;
MODEL MISSING:
            [u2-u4@-15];
            u2 ON u1@30;
            u3 ON u1-u2@30;
            u4 ON u1-u3@30;
ANALYSIS:   ESTIMATOR = MLR;
MODEL:      [u1$1*2 u2$1*1.5 u3$1*1 u4$1*1];
            f BY u1-u4@1;
            f ON x*.5;
            f@0;
OUTPUT:     TECH8 TECH9;
```

In this example, data are generated and analyzed for a discrete-time survival model like the one shown in Example 6.19. Maximum likelihood estimation with discrete-time survival analysis for a non-repeatable event requires that the event history indicators for an individual are scored as missing after an event has occurred (Muthén & Masyn, 2005). This is accomplished using the MODEL MISSING command.

The MISSING option in the MONTECARLO command is used to identify the dependent variables in the data generation model for which missing data will be generated. The MODEL MISSING command is used to provide information about the population parameter values for the missing data model to be used in the generation of data. The

MODEL MISSING command specifies a logistic regression model for a set of binary dependent variables that represent not missing (scored as 0) and missing (scored as 1) for the dependent variables in the data generation model. The binary missing data indicators have the same names as the dependent variables in the data generation model. The first statement in the MODEL MISSING command defines the intercepts in the logistic regressions for the binary dependent variables u2, u3, and u4. If the covariates predicting missingness all have values of zero, the logistic regression intercept value of -15 corresponds to a probability of zero of having missing data on the dependent variables. The variable u1 has no missing values. The first ON statement describes the regression of the missing value indicator u2 on the event-history variable u1 where the logistic regression coefficient is fixed at 30 indicating that observations with the value one on the event-history variable u1 result in a logit value 15 for the missing value indicator u2 indicating that the probability that the event-history variable u2 is missing is one. The second ON statement describes the regression of the missing value indicator u3 on the event-history variables u1 and u2 where the logistic regression coefficients are fixed at 30 indicating that observations with the value one on either or both of the event-history variables u1 and u2 result in a logit value of at least 15 for the missing value indicator u3 indicating that the probability that the event-history variable u3 is missing is one. The third ON statement describes the regression of the missing value indicator u4 on the event-history variables u1, u2, and u3 where the logistic regression coefficients are fixed at 30 indicating that observations with the value one on one or more of the event-history variables u1, u2, and u3 result in a logit value of at least 15 for the missing value indicator u4 indicating that the probability that the event-history variable u4 is missing is one. An explanation of the other commands can be found in Examples 12.1 and 12.3.

EXAMPLE 12.9: MONTE CARLO SIMULATION STUDY FOR A TWO-PART (SEMICONTINUOUS) GROWTH MODEL FOR A CONTINUOUS OUTCOME

```
TITLE:      this is an example of a Monte Carlo
            simulation study for a two-part
            (semicontinuous) growth model for a
            continuous outcome
MONTECARLO:
            NAMES = u1-u4 y1-y4;
            NOBSERVATIONS = 500;
            NREPS = 100;
            GENERATE = u1-u4(1);
            MISSING = y1-y4;
            CATEGORICAL = u1-u4;
MODEL POPULATION:
            iu su | u1@0 u2@1 u3@2 u4@3;
            [u1$1-u4$1*-.5] (1);
            [iu@0 su*.85];
            iu*1.45;
            iy sy | y1@0 y2@1 y3@2 y4@3;
            [y1-y4@0];
            y1-y4*.5;
            [iy*.5 sy*1];
            iy*1;
            sy*.2;
            iy WITH sy*.1;
            iu WITH iy*0.9;
MODEL MISSING:
            [y1-y4@15];
            y1 ON u1@-30;
            y2 ON u2@-30;
            y3 ON u3@-30;
            y4 ON u4@-30;
ANALYSIS:  ESTIMATOR = MLR;
```

```
MODEL:      iu su | u1@0 u2@1 u3@2 u4@3;
            [u1$1-u4$1*-.5] (1);
            [iu@0 su*.85];
            iu*1.45;
            su@0;
            iy sy | y1@0 y2@1 y3@2 y4@3;
            [y1-y4@0];
            y1-y4*.5;
            [iy*.5 sy*1];
            iy*1;
            sy*.2;
            iy WITH sy*.1;
            iu WITH iy*0.9;
            iu WITH sy@0;
OUTPUT:     TECH8;
```

In this example, data are generated and analyzed for a two-part
(semicontinuous) growth model for a continuous outcome like the one
shown in Example 6.16. If these data are saved for subsequent two-part
analysis using the DATA TWOPART command, an adjustment to the
saved data must be made using the DEFINE command as part of the
analysis. If the values of the continuous outcomes y are not 999 which is
the value used as the missing data flag in the saved data, the exponential
function must be applied to the continuous variables. After that
transformation, the value 999 must be changed to zero for the continuous
variables. This represents the floor of the scale.

The MISSING option in the MONTECARLO command is used to
identify the dependent variables in the data generation model for which
missing data will be generated. The MODEL MISSING command is
used to provide information about the population parameter values for
the missing data model to be used in the generation of data. The
MODEL MISSING command specifies a logistic regression model for a
set of binary dependent variables that represent not missing (scored as 0)
and missing (scored as 1) for the dependent variables in the data
generation model. The binary missing data indicators have the same
names as the dependent variables in the data generation model. The first
statement in the MODEL MISSING command defines the intercepts in
the logistic regressions for the binary dependent variables y1, y2, y3, and
y4. If the covariates predicting missingness all have values of zero, the
logistic regression intercept value of 15 corresponds to a probability of
one of having missing data on the dependent variables. The four ON
statements describe the regressions of the missing value indicators y1,
y2, y3, and y4 on the binary outcomes u1, u2, u3, and u4 where the

logistic regression coefficient is fixed at -30. This results in observations with the value one on u1, u2, u3, and u4 giving logit values -15 for the binary missing data indicators. A logit value -15 implies that the probability that the continuous outcomes y are missing is zero. An explanation of the other commands can be found in Examples 12.1 and 12.3.

EXAMPLE 12.10: MONTE CARLO SIMULATION STUDY FOR A TWO-LEVEL CONTINUOUS-TIME SURVIVAL ANALYSIS USING COX REGRESSION WITH A RANDOM INTERCEPT AND A FRAILTY

```
TITLE:      this is an example of a Monte Carlo
            simulation study for a two-level
            continuous-time survival analysis using
            Cox regression with a random intercept and
            a frailty
MONTECARLO:
            NAMES = t x w;
            NOBSERVATIONS = 1000;
            NREPS = 100;
            GENERATE = t(s 20*1);
            NCSIZES = 3;
            CSIZES = 40 (5) 50 (10) 20 (15);
            HAZARDC = t (.5);
            SURVIVAL = t (ALL);
            WITHIN = x;
            BETWEEN = w;
MODEL POPULATION:
            %WITHIN%
            x@1;
            t ON x*.5;
            %BETWEEN%
            w@1;
            [t#1-t#21*1];
            t ON w*.2;
            t*0.5;
ANALYSIS:   TYPE = TWOLEVEL;
            BASEHAZARD = OFF;
MODEL:      %WITHIN%
            t ON x*.5;
            %BETWEEN%
            t ON w*.2;
            t*0.5;
```

In this example, data are generated and analyzed for the two-level continuous-time survival analysis using Cox regression with a random intercept and a frailty shown in Example 9.16. Monte Carlo simulation of continuous-time survival models is described in Asparouhov et al. (2006).

The GENERATE option is used to specify the scale of the dependent variables for data generation. In this example, the dependent variable t is a time-to-event variable. The numbers in parentheses specify that twenty time intervals of length one will be used for data generation. The HAZARDC option is used to specify the hazard for the censoring process in continuous-time survival analysis when time-to-event variables are generated. This information is used to create a censoring indicator variable where zero is not censored and one is right censored. A hazard for censoring of .5 is specified for the time-to-event variable t by placing the number .5 in parentheses following the variable name. The SURVIVAL option is used to identify the analysis variables that contain information about time to event and to provide information about the time intervals in the baseline hazard function to be used in the analysis. The keyword ALL is used if the time intervals are taken from the data.

The ANALYSIS command is used to describe the technical details of the analysis. By selecting TYPE=TWOLEVEL, a multilevel model will be estimated. The BASEHAZARD option is used with continuous-time survival analysis to specify if a non-parametric or a parametric baseline hazard function is used in the estimation of the model. The default is OFF which uses the non-parametric baseline hazard function.

The MODEL command is used to describe the analysis model as in regular analyses. In the within part of the model, the ON statement describes the loglinear regression of the time-to-event variable t on the covariate x. In the between part of the model, the ON statement describes the linear regression of the random intercept of the time-to-event variable t on the covariate w. The residual variance of t is estimated and represents a frailty parameter. A detailed explanation of the MODEL command can be found in Examples 12.1 and 12.4.

EXAMPLE 12.11: MONTE CARLO SIMULATION STUDY FOR A TWO-LEVEL MEDIATION MODEL WITH RANDOM SLOPES

```
TITLE:      this is an example of a Monte Carlo
            simulation study for a two-level mediation
            model with random slopes
MONTECARLO:
            NAMES ARE y m x;
            WITHIN = x;
            NOBSERVATIONS = 1000;
            NCSIZES = 1;
            CSIZES = 100 (10);
            NREP = 100;
ANALYSIS:   TYPE = TWOLEVEL RANDOM;
MODEL POPULATION:
            %WITHIN%
            x@1;
            c | y ON x;
            b | y ON m;
            a | m ON x;
            m*1; y*1;
            %BETWEEN%
            y WITH m*0.1 b*0.1 a*0.1 c*0.1;
            m WITH b*0.1 a*0.1 c*0.1;
            a WITH b*0.1 (cab);
            a WITH c*0.1;
            b WITH c*0.1;
            y*1 m*1 a*1 b*1 c*1;
            [a*0.4] (ma);
            [b*0.5] (mb);
            [c*0.6];
```

```
MODEL:
        %WITHIN%
        c | y ON x;
        b | y ON m;
        a | m ON x;
        m*1; y*1;
        %BETWEEN%
        y WITH m*0.1 b*0.1 a*0.1 c*0.1;
        m WITH b*0.1 a*0.1 c*0.1;
        a WITH b*0.1 (cab);
        a WITH c*0.1;
        b WITH c*0.1;
        y*1 m*1 a*1 b*1 c*1;
        [a*0.4] (ma);
        [b*0.5] (mb);
        [c*0.6];
MODEL CONSTRAINT:
        NEW(m*0.3);
        m=ma*mb+cab;
```

In this example, data for a two-level mediation model with a random slope are generated and analyzed. For related modeling see Bauer et al. (2006).

The TYPE option is used to describe the type of analysis that is to be performed. By selecting TWOLEVEL RANDOM, a multilevel model with random intercepts and random slopes will be estimated. In the MODEL command, the | statement is used to name and define the random slopes c, b, and a. The random intercept uses the name of the dependent variables c, b, and a. The ON statements on the right-hand side of the | statements describe the linear regressions that have a random slope.

The label cab is assigned to the covariance between the random slopes a and b. The labels ma and mb are assigned to the means of the random slopes a and b. These labels are used in the MODEL CONSTRAINT command. The MODEL CONSTRAINT command is used to define linear and non-linear constraints on the parameters in the model. In the MODEL CONSTRAINT command, the NEW option is used to introduce a new parameter that is not part of the MODEL command. The new parameter m is the indirect effect of the covariate x on the outcome y. The two outcomes y and m can also be categorical. For a discussion of indirect effects when the outcome y is categorical, see MacKinnon et al. (2007).

The default estimator for this type of analysis is maximum likelihood with robust standard errors. An explanation of the other commands can be found in Examples 12.1 and 12.4.

EXAMPLE 12.12: MONTE CARLO SIMULATION STUDY FOR A MULTIPLE GROUP EFA WITH CONTINUOUS FACTOR INDICATORS WITH MEASUREMENT INVARIANCE OF INTERCEPTS AND FACTOR LOADINGS

```
TITLE:     this is an example of a Monte Carlo
           simulation study for a multiple group EFA
           with continuous factor indicators with
           measurement invariance of intercepts and
           factor loadings
MONTECARLO:
           NAMES ARE y1-y10;
           NOBSERVATIONS = 500 500;
           NREPS = 1;
           NGROUPS = 2;
MODEL POPULATION:
           f1 BY y1-y5*.8 y6-y10*0;
           f2 BY y1-y5*0 y6-y10*.8;
           f1-f2@1;
           f1 WITH f2*.5;
           y1-y10*1;
           [y1-y10*1];
           [f1-f2@0];
MODEL POPULATION-g2:
           f1*1.5 f2*2;
           f1 WITH f2*1;
           y1-y10*2;
           [f1*.5 f2*.8];
MODEL:     f1 BY y1-y5*.8 y6-y10*0 (*1);
           f2 BY y1-y5*0 y6-y10*.8 (*1);
           f1-f2@1;
           f1 WITH f2*.5;
           y1-y10*1;
           [y1-y10*1]; [f1-f2@0];
MODEL g2:  f1*1.5 f2*2;
           f1 WITH f2*1;
           y1-y10*2; [f1*.5 f2*.8];
OUTPUT:    TECH9;
```

In this example, data are generated and analyzed according to a multiple group EFA model with continuous factor indicators with measurement

invariance across groups of intercepts and factor loadings. This model is described in Example 5.27. The NOBSERVATIONS option specifies the number of observations for each group. The NGROUPS option specifies the number of groups. In this study data for two groups of 500 observations are generated and analyzed. One difference between the MODEL command when EFA factors are involved rather than CFA factors is that the values given using the asterisk (*) are used only for coverage. Starting values are not allowed for the factor loading and factor covariance matrices for EFA factors. An explanation of the other commands can be found in Example 12.1 and Example 5.27.