How To Use A Monte Carlo Study To Decide On Sample Size and Determine Power

Linda K. Muthén
Muthén & Muthén
11965 Venice Blvd., Suite 407
Los Angeles, CA 90066
Telephone: (310) 391-9971
Fax: (310) 391-8971
muthen@statmodel.com

Bengt Muthén
University of California, Los Angeles
Graduate School of Education & Information Studies
2023 Moore Hall, Mailbox 951521
Los Angeles, CA 90095-1521
Telephone: (310) 206-1226
Fax: (310) 206-6293
bmuthen@ucla.edu

Running head: SAMPLE SIZE AND POWER

April 9, 2002

ABSTRACT

A common question asked by researchers is, “What sample size do I need for my study?” Over the years, several rules of thumb have been proposed. In reality there is no rule of thumb that applies to all situations. The sample size needed for a study depends on many factors including the size of the model, distribution of the variables, amount of missing data, reliability of the variables, and strength of the relationships among the variables. The purpose of this paper is to demonstrate how substantive researchers can use a Monte Carlo study to decide on sample size and determine power. Two models are used as examples, a confirmatory factor analysis (CFA) model and a growth model. The analyses are carried out using the Mplus program (Muthén & Muthén, 1998).

A common question asked by researchers is, "What sample size do I need for my study?" Over the years, several rules of thumb have been proposed such as 5-10 observations per parameter, 50 observations per variable, no less than 100, and so on. In reality there is no rule of thumb that applies to all situations. The sample size needed for a study depends on many factors including the size of the model, distribution of the variables, amount of missing data, reliability of the variables, and strength of the relationships among the variables. Although parameter estimates frequently have small bias, standard errors are more sensitive. Standard errors may be overestimated or underestimated depending on the situation. This affects the estimation of confidence intervals also referred to as coverage. If standard errors are overestimated, significant effects may be missed. If they are underestimated, significant effects may be overstated. Another issue that needs to be considered when deciding on sample size is power. A sample may be large enough for unbiased parameter estimates, unbiased standard errors, and good coverage, but it may not be large enough to detect an important effect in the model.

The purpose of this paper is to demonstrate how substantive researchers can use a Monte Carlo study to decide on sample size and determine power. Two models are used as examples, a confirmatory factor analysis (CFA) model and a growth model. The analyses are carried out using the Mplus program (Muthén & Muthén, 1998) which has extensive Monte Carlo facilities. Data generation using Mplus can include normal data, non-normal data, missing data, clustering, and mixtures of populations. Analysis models can include any of the models available in Mplus. Data generation and analysis models do not need to be the same.

This paper focuses on parameter estimates, standard errors, coverage, and power assuming correctly specified models. Mis-specified models can also be studied in the Mplus Monte Carlo framework, but are not included here. Also, it should be noted that Monte Carlo studies are useful for evaluating the performance of model fit indices, but this use is not considered in the paper.

METHOD

A common use of Monte Carlo studies is for methodological investigations of the performance of statistical estimators under various conditions. In these studies, data are generated and models are estimated, sometimes using more than one estimator. The performance of an estimator is judged by studying parameter estimate bias, standard error bias, and coverage. A less common use of Monte Carlo studies is to decide on sample size and determine power in the design of substantive studies. This use is the focus of the paper.

MONTE CARLO STUDY

In Monte Carlo studies, data are generated from a population with hypothesized parameter values. A large number of samples are drawn, and a model is estimated for each sample. Parameter values and standard errors are averaged over the samples. The

following criteria are examined: parameter estimate bias, standard error bias, and coverage.

Several decisions need to be made to carry out a Monte Carlo study. The first is the choice of the model to be studied. This choice is driven by the research question being asked. Once the model is chosen, population values for each parameter of the model must be selected. These values can be obtained from theory or previous research. Estimates from previous studies are often the best estimates available for population values in the Monte Carlo study.

Technical considerations in the Monte Carlo study are the number of samples to be drawn and the seed. The number of samples to be drawn (replications) can be thought of as the sample size for the Monte Carlo study. The number of replications should be increased until stability of the results is achieved. In this study, 10,000 replications are used for each analysis to insure that stability has been reached. The value of the seed determines the starting point for the random draws of the samples. More than one seed should be used, and the results for the different seeds should be checked for stability.

MODELS TO BE STUDIED

A CFA model and a growth model were selected for study. These models were chosen because they are often used in practice and are sufficiently different from each other. CFA models are typically cross-sectional and have only a covariance structure. The growth model is longitudinal and has both a mean and covariance structure.

Confirmatory Factor Analysis Model

The CFA model that is studied has two factors, each of which has five continuous factor indicators. The CFA model has 31 free parameters and 24 degrees of freedom. A diagram of the CFA model is shown in Figure 1. Data are generated using the following population values. The factor loadings are 0.8. The residual variances of the factor indicators are 0.36. Factor variances are fixed to one to set the metric of the factors. The

Insert Figure 1 Here

factor correlation is 0.25. All factor loadings are free. These population values are chosen so that the variances of the factor indicators are one which makes the parameter values more easily interpretable. The population values result in a reliability of 0.64 for each factor indicator. Reliability is calculated as the ratio of the variance of the factor indicator explained by the factor to the total variance of the factor indicator using the following formula,

$$(1) \qquad \lambda^2 \psi / (\lambda^2 \psi + \theta) ,$$

where λ is the factor loading, ψ is the factor variance, and θ is the residual variance.

The focus of the power investigation in the CFA model is the factor correlation. This parameter is of particular interest because it represents the correlation between the two constructs unattenuated by measurement error. The CFA model can also be thought of as a longitudinal model with two measurement occasions so that the last five indicators are repeated measures of the first five indicators. In this case, the factor correlation can be seen as a measure of stability of the construct over time.

The CFA model is examined under four conditions: (1) normally distributed continuous factor indicators without missing data, (2) normally distributed continuous factor indicators with missing data, (3) non-normal continuous factor indicators without missing data, and (4) non-normal continuous factor indicators with missing data.

Missing Data

In the analyses with missing data, the data are generated such that all subjects have data on y1, y2, y3, y4, and y5 and 50 percent of the subjects have data on y6, y7, y8, y9, and y10. The patterns of missing data should be specified to reflect missing data patterns seen in practice. For example, the percent of missing data can increase in relation to the number of questions in a survey to reflect the likelihood that subjects become tired toward the end of a survey and start skipping questions. Or the percent of missing data can increase over time reflecting the likelihood that people will drop out of a study. If a study is designed such that some subjects receive only a subset of the items on a survey or are measured only at certain ages, this can also be reflected in the generation of data. The way missing data are generated for the CFA model is an example of missing completely at random (MCAR; Little & Rubin, 1987).

Non-Normal Data

In the analyses with non-normal data, the data are created using a mixture of two normal subpopulations or classes of individuals. Normal data are generated for two classes that have different means and variances for the factor indicators. The combined data are analyzed as though they come from a single population. To maintain a similarity between the CFA models without and with missing data, the parameter values for the factor indicators are chosen so that their reliabilities are 0.64 using equation (1).

The first step is to generate data for two classes such that the combination of the data from the two classes has the desired skewness and kurtosis. This is done by allowing one of the classes to represent an outlying group of individuals that has different means and variances for the factor indicators. The choice of the proportion of individuals in the two classes also affects skewness and kurtosis. To insure that the model for the combined data is a correctly specified CFA model, skewness and kurtosis in the factor indicators is achieved by choosing different means and variances for the factors, not by manipulating the means and variances of the factor indicators.

For the CFA model with non-normal data, Class 1, the outlier class, contains 12 percent of the subjects and Class 2 contains the remaining 88 percent. Only the factor indicators

for the second factor are non-normal. Therefore, the Class 1 mean for the second factor is chosen to be 15 and the variance 5 as compared to the Class 2 mean and variance of zero and one. The resulting population univariate skewness for variables y6 through y10 is 1.2. The resulting population univariate kurtosis for variables y6 through y10 ranges from 1.5 to 1.6.

The second step is to run the analysis with one replication and a large sample to obtain approximate population values for the one class model. In this paper, a sample size of 100,000 is used. Given that factor indicator reliabilities of 0.64 are desired, the third step is to solve for the population residual variances for the factor indicators of the second factor using equation (1) and use those values as the population values for data generation.

Growth Model

Two growth models are studied. Both are linear growth models with equidistant time points for four continuous outcomes. One has a covariate influencing the intercept and slope growth factors. The growth model without a covariate has 9 free parameters and 5 degrees of freedom. The growth model with a covariate has 11 free parameters and 7 degrees of freedom. Figure 2 shows the diagram for the growth model with the covariate. Data are generated using the following population values. For the growth model without a covariate, the mean of the intercept growth factor is 0.0 and the mean of the slope growth factor is 0.2. The variance of the intercept growth factor is 0.5 and the variance

Insert Figure 2 Here

of the slope growth factor is 0.1, reflecting a commonly seen variance ratio. The covariance between the intercept and slope growth factors is zero. The residual variances of the continuous outcomes are 0.5. This results in R-square values of 0.50 for y1, 0.55 for y2, 0.64 for y3, and 0.74 for y4 using the following formula,

(2) $$\text{R-square}\,(y_t) = (\psi_i + x_t^2\,\psi_s + 2\,x_t\,\psi_{is})/(\psi_i + x_t^2\,\psi_s + 2\,x_t\,\psi_{is} + \theta_t),$$

where ψ_i is the variance of the intercept growth factor, x_t is the time score at time t, ψ_s is the variance of the slope growth factor, ψ_{is} is the covariance between intercept and slope growth factors (set at zero in this case), and θ_t is the residual variance for the outcome at time t. Here the x_t time scores are chosen as 0, 1, 2, and 3.

In the growth model with a covariate, the intercept and slope growth factors are regressed on a dichotomous covariate with a 50/50 split giving the covariate a mean of 0.5 and a variance of 0.25. This covariate can be thought of as a treatment or gender dummy variable. For the intercept growth factor, the regression coefficient is 0.5. The residual variance for the intercept growth factor is chosen as 0.25. This corresponds to an R-square value of 0.20 for the intercept growth factor.

The focus of the power investigation in the growth model is the regression coefficient in the regression of the slope growth factor on the covariate. This parameter is selected because across-group differences in development over time are the focus of many longitudinal studies. Regression coefficient values of 0.2 and 0.1 are chosen to study different effect sizes. A regression coefficient of 0.2 has an effect size of 0.63 reflecting a medium effect (Cohen, 1969). A slope of 0.1 has an effect size of 0.32 reflecting a small effect. Here effect size is computed as the ratio of the difference in the slope means for the two values of the covariate divided by the standard deviation of the slope growth factor. The residual variance for the slope growth factor is chosen as 0.09. This corresponds to an R-square value of 0.10 for the slope growth factor when the regression coefficient is 0.2 and an R-square of 0.03 when the regression coefficient is 0.1. Values as low as these are commonly seen in the prediction of the slope growth factor.

The growth model is examined under five conditions: (1) normally distributed continuous outcomes without missing data without a covariate, (2) normally distributed continuous outcomes without missing data with a covariate that has a regression coefficient of 0.2 for the slope growth factor, (3) normally distributed continuous outcomes with missing data with a covariate that has a regression coefficient of 0.2 for the slope growth factor, (4) normally distributed continuous outcomes without missing data with a covariate that has a regression coefficient of 0.1 for the slope growth factor, and (5) normally distributed continuous outcomes with missing data with a covariate that has a regression coefficient of 0.1 for the slope growth factor.

Missing Data

In the analyses with missing data, the data are generated to reflect an increase in missing data over time due to attrition. For the second through the fourth time points, the probability of missing data is influenced by the covariate, while the first time point has data missing completely at random (MCAR). For the covariate value of zero, the first measurement occasion has 12 percent missing on the outcome, the second has 18 percent missing, the third has 27 percent missing, and the fourth has 50 percent missing. For the covariate value of one, the first measurement occasion has 12 percent missing on the outcome, the second has 38 percent missing, the third has 50 percent missing, and the fourth has 73 percent missing. The way missing data are generated for the growth model is an example of missing at random (MAR; Little & Rubin, 1987).

MODEL ESTIMATION

Model estimation is carried out in all cases by maximum likelihood under the assumption of normality. For models with non-normal data, standard errors are computed using a non-normality robust sandwich estimator. All analyses are done using the Mplus program. All Mplus inputs used for the paper are included in Appendix 1 and are available at www.statmodel.com. Complete outputs are also available at this website.

STRATEGY FOR DECIDING ON SAMPLE SIZE

Several criteria are examined to determine sample size. The first criterion is that parameter and standard error biases do not exceed 10 percent for any parameter in the model. The second criterion is that the standard error bias for the parameter for which power is being assessed does not exceed 5 percent. The third criterion is that coverage remains between 0.91 and 0.98. Once these three conditions are satisfied, the sample size is chosen to keep power close to 0.80. The value of 0.80 is used because it is a commonly accepted value for sufficient power.

Appendix 2 shows partial output from the Mplus analysis for the CFA model with normally distributed continuous factor indicators without missing data. All outputs from the analyses in this paper are available at the website www.statmodel.com. Following is a description of how the information in the output is used to evaluate the criteria discussed above.

Parameter bias is evaluated using the information in columns one and two of the output. The column labeled Starting gives the population parameter values. The column labeled Average gives the parameter estimate average over the replications of the Monte Carlo study. For example, the first number in column 2, 0.7963, is the average of the factor loading estimates for y1 over 10,000 replications. To determine its bias, subtract the population value of 0.8 from this number and divide it by the population value of 0.8. This results in a bias of -0.005 which is negligible.

Standard error bias is evaluated using the information in columns three and four of the output. The column labeled Std. Dev. gives the standard deviation of each parameter estimate over the replications of the Monte Carlo study. This is considered to be the population standard error when the number of replications is large. The column labeled S.E. Average gives the average of the estimated standard errors for each parameter estimate over the replications of the Monte Carlo study. Standard error bias is calculated in the same way as parameter estimate bias as described above.

Coverage is evaluated using the information in column 6 of the output labeled 95% Cover. It gives the proportion of replications for which the 95% confidence interval contains the true parameter value.

Power is evaluated using the information in column 7 of the output labeled % Sig Coeff. This column gives the proportion of replications for which the null hypothesis that a parameter is equal to zero is rejected for each parameter 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, 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.

FINDINGS

CONFIRMATORY FACTOR ANALYSIS MODEL

The results of the four CFA analyses are found in Table 1. For the simplest CFA model with normally distributed continuous factor indicators and no missing data, a sample size of 150 is needed for power of 0.81 to reject the hypothesis that the factor correlation is zero. By adding the complication of missing data, a sample size of 175 is required for power of 0.81. Considering the CFA model with non-normal factor indicators without missing data, a sample size of 265 is needed for a power of 0.80. Adding the complication of missing data results in the need for a sample size of 315 for power of 0.81.

Insert Table 1 Here

GROWTH MODEL

The results of the five growth model analyses are found in Table 2. For the simplest growth model without missing data and without a covariate, a sample size of 40 is needed for power of 0.81 to reject the hypothesis that the mean of the slope growth factor is zero. By adding a dichotomous covariate with population regression coefficient of 0.2 for the regression of the slope growth factor on the covariate, the sample size requirement to reject the hypothesis that the regression coefficient is zero rises to 150 for a power of 0.81. By adding the complication of missing data, the sample size requirement increases to 250 for a power of 0.80. By eliminating the missing data complication and changing the population value of the regression coefficient to 0.1, the sample size requirement is 600 for a power of 0.80. By adding the complication of missing data to the model with a regression coefficient of 0.1, the samples size requirement rises to 1025 for a power of 0.80.

Insert Table 2 Here

DISCUSSION

This paper demonstrated the use of a Monte Carlo study for the purpose of deciding on sample size and determining power. A CFA and a growth model were considered.

For the CFA model, the influences of non-normality and missing data on sample size requirements were studied. Sample size requirements were found to be influenced more by non-normality than missing data, at least in this situation where data are missing completely at random. For both normal and non-normal data, adding the complication of missing data increased the sample size requirement by approximately 18 percent. Having both non-normality and missing data approximately doubled the sample size requirement.

For the growth model, the influence of missing data, a covariate, and regression coefficient size on sample size requirements were studied. It was found that the largest impact on the sample size requirement came from including a small regression coefficient for the covariate in the model. Reducing the population value of the regression coefficient from 0.2 to 0.1 increased the sample size requirement approximately four times both with and without missing data. This reflected a change in effect size from medium to small. Including missing data in the model increased the sample size requirement by a factor of approximately 1.7 for both effect sizes.

The results in this paper support the fact that sample size requirements depend strongly on many factors. As an example, the sample size requirement of 600 for detecting a small effect size in the growth model is high in contrast to the sample size requirement of 265 for detecting a small factor correlation in the CFA model.

The paper demonstrated how substantive researchers can use a Monte Carlo study to decide on sample size and determine power. Two models were considered and a strategy for deciding on sample size was described. Many variations of the models and strategy described in the paper can also be considered. Variations of the CFA model that can be considered are factor cross-loadings and/or residual covariances. In addition, the number of factors and the number of factor indicators can be varied. Variations of the growth model that can be considered are different choices of the R-square value for the slope growth factor and the continuous outcomes, residual covariances, free time scores, quadratic models, and piecewise models. In addition, the number of time points can be varied. Also, if a researcher is interested in power for only one parameter, it is not necessary to have the strict bias requirements for all parameters in the model as suggested in the strategy of this paper.

In addition to the models and data complications included in this paper, Monte Carlo studies in Mplus can include investigations of sample size and power in situations with cluster samples (hierarchical data) and mixtures of unobserved subpopulations. This allows studies of sample size and power for multilevel CFA models, 3-level growth models, factor mixture models, and growth mixture models. It is important to investigate the reduction in power due to cluster sampling and due to considering small subpopulations in mixture models.

REFERENCES

Cohen, J. (1969). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Little, R.J., & Rubin, D.B. (1987). Statistical analysis with missing data. New York: John Wiley & Sons.

Muthén, L.K. and Muthén, B.O. (1998). Mplus user's guide. Los Angeles, CA: Muthén & Muthén.

ACKNOWLEDGEMENTS

Preparation of this paper was supported by Grant K02 AA 00230 and SBIR Contract N44AA92009 both from the National Institute on Alcohol Abuse and Alcoholism. This paper is based on Mplus Web Note No. 1, Using Mplus Monte Carlo Simulations In Practice: A Note On Assessing Estimation Quality and Power in Latent Variable Models, authored by Bengt Muthén and Mplus Web Note No. 2, Using Mplus Monte Carlo Simulations In Practice: A Note On Non-Normal Missing Data In Latent Variable Models, authored by Bengt Muthén and Tihomir Asparouhov. These notes can be found at www.statmodel.com.

APPENDIX 1

Appendix 1 contains the Mplus input files for the nine analyses in the paper. Following is a brief description of the Mplus commands. Details about the input language can be found in the Mplus User's Guide (Muthén & Muthén, 1998). The TITLE command provides a title for the output. The MONTECARLO command describes the technical details of the Monte Carlo study. The ANALYSIS command provides information about the type of analysis to be performed. The MODEL MONTECARLO command is used to provide the population parameter values to be used in data generation. The MODEL command describes the model to be estimated. The OUTPUT command is used to request extra output.

Mplus Input File For The CFA Model With Normally Distributed Continuous Factor Indicators Without Missing Data

```
TITLE:          cfa1.inp normal, no missing
MONTECARLO:     NAMES ARE y1-y10;
                NOBSERVATIONS = 150;
                NREPS = 10000;
                SEED = 53487;
                NCLASSES = 1;
                GCLASSES = 1;
                SAVE = cfa1.sav;
ANALYSIS:       TYPE = MIXTURE;
                ESTIMATOR = ML;
MODEL MONTECARLO:
                %OVERALL%
                f1 BY y1-y5*.8;
                f2 BY y6-y10*.8;
                f1@1 f2@1;
                y1-y10*.36;
                f1 WITH f2*.25;
MODEL:
                %OVERALL%
                f1 BY y1-y5*.8;
                f2 BY y6-y10*.8;
                f1@1 f2@1;
                y1-y10*.36;
                f1 WITH f2*.25;
OUTPUT:         TECH9;
```

Mplus Input File For The CFA Model With Normally Distributed Continuous Factor Indicators With Missing Data

```
TITLE:            cfa2.inp  normal, missing
MONTECARLO:       NAMES ARE y1-y10;
                  NOBSERVATIONS = 175;
                  NREPS = 10000;
                  SEED = 53487;
                  NCLASSES = 1;
                  GCLASSES = 1;
```

```
                  PATMISS = y6 (.5) y7 (.5) y8 (.5) y9 (.5) y10 (.5);
                  PATPROB = 1;
                  SAVE = cfa2.sav;
ANALYSIS:         TYPE = MIXTURE MISSING;
                  ESTIMATOR = ML;
MODEL MONTECARLO:
                  %OVERALL%
                  f1 BY y1-y5*.8;
                  f2 BY y6-y10*.8;
                  f1@1 f2@1;
                  y1-y10*.36;
                  f1 WITH f2*.25;
MODEL:
                  %OVERALL%
                  f1 BY y1-y5*.8;
                  f2 BY y6-y10*.8;
                  f1@1 f2@1;
                  y1-y10*.36;
                  f1 WITH f2*.25;
OUTPUT:           PATTERNS TECH9;
```

Mplus Input File For The CFA Model With Non-Normal Continuous Factor Indicators Without Missing Data

```
TITLE:            cfa3.inp  non-normal, no missing
MONTECARLO:       NAMES ARE y1-y10;
                  NOBSERVATIONS = 265;
                  NREPS = 10000;
                  SEED = 53487;
                  NCLASSES = 1;
                  GCLASSES = 2;
                  SAVE = cfa3.sav;
ANALYSIS:         TYPE = MIXTURE;
                  ESTIMATOR = MLR;
MODEL MONTECARLO:
                  %OVERALL%
                  f1 BY y1-y5*.8;
                  f2 BY y6-y10*.8;
                  f1@1 f2@1;
                  y1-y5*.36 y6-y10*9;
                  f1 WITH f2*.95;
                  [C#1@-2];

                  %C#1%

                  [f1@0 f2@15];
                  f1@1 f2@5;

                  %C#2%

                  [f1@0 f2@0];
                  f1@1 f2@1;
MODEL:
                  %OVERALL%
                  f1 BY y1-y5*.8;
```

```
                  f2 BY y6-y10*4;
                  f1@1 f2@1;
                  y1-y5*.36 y6-y10*9;
                  f1 WITH f2*.20;

                  [y6-y10*1.42];
OUTPUT:           TECH9;
```

Mplus Input File For The CFA Model With Non-Normal Continuous Factor Indicators With Missing Data

```
TITLE:            cfa4.inp  non-normal, missing
MONTECARLO:       NAMES ARE y1-y10;
                  NOBSERVATIONS = 315;
                  NREPS = 10000;
                  SEED = 53487;
                  NCLASSES = 1;
                  GCLASSES = 2;
                  PATMISS = y6 (.5) y7 (.5) y8 (.5) y9(.5) y10 (.5);
                  PATPROB = 1;
                  SAVE = cfa4.sav;
ANALYSIS:         TYPE = MIXTURE MISSING;
                  ESTIMATOR = MLR;
MODEL MONTECARLO:
                  %OVERALL%
                  f1 BY y1-y5*.8;
                  f2 BY y6-y10*.8;
                  f1@1 f2@1;
                  y1-y5*.36 y6-y10*9;
                  f1 WITH f2*.95;
                  [C#1@-2];

                  %C#1%

                  [f1@0 f2@15];
                  f1@1 f2@5;

                  %C#2%

                  [f1@0 f2@0];
                  f1@1 f2@1;
MODEL:
                  %OVERALL%
                  f1 BY y1-y5*.8;
                  f2 BY y6-y10*4;
                  f1@1 f2@1;
                  y1-y5*.36 y6-y10*9;
                  f1 WITH f2*.20;

                  [y6-y10*1.42];
OUTPUT:           PATTERNS TECH9;
```

Mplus Input File For The Growth Model With Normally Distributed Continuous Outcomes Without Missing Data Without A Covariate

```
TITLE:            growth1.inp  normal, no covariate, no missing
MONTECARLO:       NAMES ARE y1-y4;
                  NOBSERVATIONS = 40;
                  NREPS = 10000;
                  SEED = 53487;
                  NCLASSES = 1;
                  GCLASSES = 1;
                  SAVE = growth1.sav;
ANALYSIS:         TYPE = MIXTURE;
                  ESTIMATOR = ML;
MODEL MONTECARLO:
                  %OVERALL%
                  i BY y1-y4@1;
                  s BY y1@0 y2@1 y3@2 y4@3;
                  [y1-y4@0];
                  [i*0 s*.2];
                  i*.5;
                  s*.1;
                  i WITH s*0;
                  y1-y4*.5;

                  %C#1%

                  [i*0 s*.2];
MODEL:
                  %OVERALL%
                  i BY y1-y4@1;
                  s BY y1@0 y2@1 y3@2 y4@3;
                  [y1-y4@0];
                  [i*0 s*.2];
                  i*.5;
                  s*.1;
                  i WITH s*0;
                  y1-y4*.5;

                  %C#1%

                  [i*0 s*.2];
OUTPUT:           TECH9;
```

Mplus Input File For The Growth Model With Normally Distributed Continuous Outcomes Without Missing Data With A Covariate That Has A Regression Coefficient Of 0.2 For The Slope Growth Factor

```
TITLE:            growth2.inp  normal, covariate, no missing
MONTECARLO:       NAMES ARE y1-y4 x;
                  CUTPOINTS = x (0);
                  NOBSERVATIONS = 150;
                  NREPS = 10000;
                  SEED = 53487;
                  NCLASSES = 1;
```

```
                  GCLASSES = 1;
                  SAVE = growth2.sav;
ANALYSIS:         TYPE = MIXTURE;
                  ESTIMATOR = ML;
MODEL MONTECARLO:
                  %OVERALL%
                  [x@0]; x@1;
                  i BY y1-y4@1;
                  s BY y1@0 y2@1 y3@2 y4@3;
                  [y1-y4@0];
                  [i*0 s*.2];
                  i*.25;
                  s*.09;
                  i WITH s*0;
                  y1-y4*.5;

                  i ON x*.5;
                  s ON x*.2;

                  %C#1%

                  [i*0 s*.2];
MODEL:
                  %OVERALL%
                  i BY y1-y4@1;
                  s BY y1@0 y2@1 y3@2 y4@3;
                  [y1-y4@0];
                  [i*0 s*.2];
                  i*.25;
                  s*.09;
                  i WITH s*0;
                  y1-y4*.5;

                  i ON x*.5;
                  s ON x*.2;

                  %C#1%

                  [i*0 s*.2];
OUTPUT:           TECH9;
```

Mplus Input File For The Growth Model With Normally Distributed Continuous Outcomes With Missing Data With A Covariate That Has A Regression Coefficient Of 0.2 For The Slope Growth Factor

```
TITLE:            growth3.inp  normal, covariate, missing
MONTECARLO:       NAMES ARE y1-y4 x;
                  CUTPOINTS = x (0);
                  NOBSERVATIONS = 250;
                  NREPS = 10000;
                  SEED = 53487;
                  NCLASSES = 1;
                  GCLASSES = 1;
```

```
                  MISSING = y1-y4;
                  SAVE = growth3.sav;
ANALYSIS:         TYPE = MIXTURE MISSING;
                  ESTIMATOR = ML;
MODEL MISSING:
                  %OVERALL%
                  [y1@-2 y2@-1.5 y3@-1 y4@0];
                  y2-y4 ON x@1;
MODEL MONTECARLO:
                  %OVERALL%
                  [x@0]; x@1;
                  i BY y1-y4@1;
                  s BY y1@0 y2@1 y3@2 y4@3;
                  [y1-y4@0];
                  [i*0 s*.2];
                  i*.25;
                  s*.09;
                  i WITH s*0;
                  y1-y4*.5;

                  i ON x*.5;
                  s ON x*.2;

                  %C#1%

                  [i*0 s*.2];
MODEL:
                  %OVERALL%
                  i BY y1-y4@1;
                  s BY y1@0 y2@1 y3@2 y4@3;
                  [y1-y4@0];
                  [i*0 s*.2];
                  i*.25;
                  s*.09;
                  i WITH s*0;
                  y1-y4*.5;

                  i ON x*.5;
                  s ON x*.2;

                  %C#1%

                  [i*0 s*.2];
OUTPUT:           TECH9;
```

Mplus Input File For The Growth Model With Normally Distributed Continuous Outcomes Without Missing Data With A Covariate That Has A Regression Coefficient Of 0.1 For The Slope Growth Factor

```
TITLE:            growth4.inp  normal, covariate, no missing
MONTECARLO:       NAMES ARE y1-y4 x;
                  CUTPOINTS = x (0);
                  NOBSERVATIONS = 600;
                  NREPS = 10000;
```

```
                  SEED = 53487;
                  NCLASSES = 1;
                  GCLASSES = 1;
                  SAVE = growth4.sav;
ANALYSIS:         TYPE = MIXTURE;
                  ESTIMATOR = ML;
MODEL MONTECARLO:
                  %OVERALL%
                  [x@0]; x@1;
                  i BY y1-y4@1;
                  s BY y1@0 y2@1 y3@2 y4@3;
                  [y1-y4@0];
                  [i*0 s*.2];
                  i*.25;
                  s*.09;
                  i WITH s*0;
                  y1-y4*.5;

                  i ON x*.5;
                  s ON x*.1;

                  %C#1%

                  [i*0 s*.2];
MODEL:
                  %OVERALL%
                  i BY y1-y4@1;
                  s BY y1@0 y2@1 y3@2 y4@3;
                  [y1-y4@0];
                  [i*0 s*.2];
                  i*.25;
                  s*.09;
                  i WITH s*0;
                  y1-y4*.5;

                  i ON x*.5;
                  s ON x*.1;

                  %C#1%

                  [i*0 s*.2];
OUTPUT:           TECH9;
```

Mplus Input File For The Growth Model With Normally Distributed Continuous Outcomes With Missing Data With A Covariate That Has A Regression Coefficient Of 0.1 For The Slope Growth Factor

```
TITLE:            growth5.inp  normal, covariate, missing
MONTECARLO:       NAMES ARE y1-y4 x;
                  CUTPOINTS = x (0);
                  NOBSERVATIONS = 1025;
                  NREPS = 10000;
                  SEED = 53487;
                  NCLASSES = 1;
```

```
                  GCLASSES = 1;
                  MISSING = y1-y4;
                  SAVE = growth5.sav;
ANALYSIS:         TYPE = MIXTURE MISSING;
                  ESTIMATOR = ML;
MODEL MISSING:
                  %OVERALL%
                  [y1@-2 y2@-1.5 y3@-1 y4@0];
                  y2-y4 on x@1;
MODEL MONTECARLO:
                  %OVERALL%
                  [x@0]; x@1;
                  i BY y1-y4@1;
                  s BY y1@0 y2@1 y3@2 y4@3;
                  [y1-y4@0];
                  [i*0 s*.2];
                  i*.25;
                  s*.09;
                  i WITH s*0;
                  y1-y4*.5;

                  i ON x*.5;
                  s ON x*.1;

                  %C#1%

                  [i*0 s*.2];
MODEL:
                  %OVERALL%
                  i BY y1-y4@1;
                  s BY y1@0 y2@1 y3@2 y4@3;
                  [y1-y4@0];
                  [i*0 s*.2];
                  i*.25;
                  s*.09;
                  i WITH s*0;
                  y1-y4*.5;

                  i ON x*.5;
                  s ON x*.1;

                  %C#1%

                  [i*0 s*.2];
OUTPUT:           TECH9;
```

APPENDIX 2

Mplus Output Excerpts For the CFA Model with Normally Distributed Continuous Factor Indicators and No Missing Data

MODEL RESULTS

		ESTIMATES		S. E.	M. S. E.	95%	% Sig
	Starting	Average	Std. Dev.	Average		Cover	Coeff
CLASS 1							
F1 BY							
Y1	0.800	0.7963	0.0707	0.0697	0.0707	0.949	1.000
Y2	0.800	0.7981	0.0712	0.0698	0.0712	0.942	1.000
Y3	0.800	0.7962	0.0708	0.0697	0.0708	0.946	1.000
Y4	0.800	0.7975	0.0708	0.0698	0.0708	0.944	1.000
Y5	0.800	0.7971	0.0704	0.0698	0.0704	0.947	1.000
F2 BY							
Y6	0.800	0.7959	0.0706	0.0697	0.0707	0.945	1.000
Y7	0.800	0.7961	0.0702	0.0697	0.0702	0.950	1.000
Y8	0.800	0.7950	0.0701	0.0697	0.0701	0.945	1.000
Y9	0.800	0.7969	0.0710	0.0698	0.0710	0.946	1.000
Y10	0.800	0.7968	0.0703	0.0698	0.0703	0.946	1.000
F1 WITH							
F2	0.250	0.2497	0.0864	0.0850	0.0864	0.942	0.812
Residual Variances							
Y1	0.360	0.3551	0.0523	0.0513	0.0523	0.934	1.000
Y2	0.360	0.3548	0.0523	0.0514	0.0523	0.933	1.000
Y3	0.360	0.3546	0.0529	0.0513	0.0529	0.929	1.000
Y4	0.360	0.3553	0.0525	0.0514	0.0525	0.931	1.000
Y5	0.360	0.3547	0.0526	0.0513	0.0527	0.934	1.000
Y6	0.360	0.3548	0.0516	0.0513	0.0516	0.939	1.000
Y7	0.360	0.3545	0.0524	0.0513	0.0524	0.929	1.000
Y8	0.360	0.3548	0.0520	0.0513	0.0521	0.934	1.000
Y9	0.360	0.3554	0.0524	0.0514	0.0524	0.935	1.000
Y10	0.360	0.3550	0.0525	0.0514	0.0526	0.934	1.000
Variances							
F1	1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
F2	1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000

TABLE 1

Sample Size Requirements For The CFA Model

	No Missing	Missing
Normal	150	175
Non-normal	265	315

TABLE 2

Sample Size Requirements For The Growth Model

	No Missing	Missing
No Covariate	40	NA
Regression Coefficient .2	150	250
Regression Coefficient .1	600	1025

Figure Captions

FIGURE 1 CFA Model.

FIGURE 2 Growth Model.

FIGURE 1

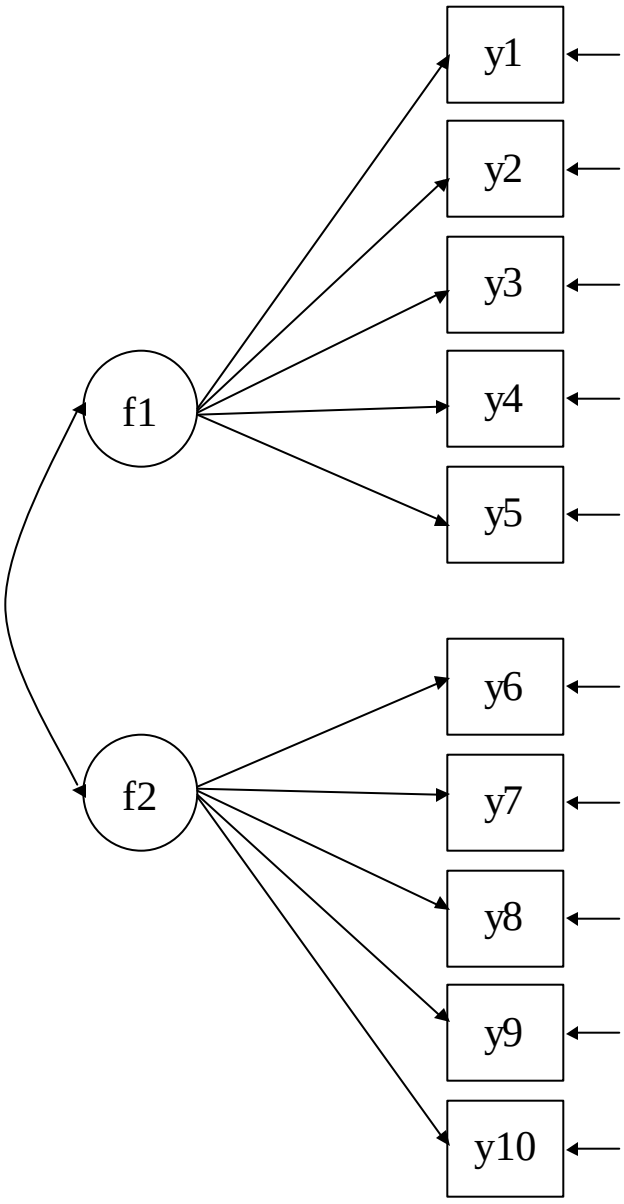

FIGURE 2