

CHAPTER 16

ANALYSIS COMMAND

In this chapter, the ANALYSIS command is discussed. The ANALYSIS command is used to describe the technical details of the analysis including the type of analysis, the statistical estimator, the parameterization of the model, and the specifics of the computational algorithms.

THE ANALYSIS COMMAND

Following are the options for the ANALYSIS command:

ANALYSIS:		
TYPE =	GENERAL; BASIC; RANDOM; COMPLEX; MIXTURE; BASIC; RANDOM; COMPLEX; TWOLEVEL; BASIC; RANDOM; MIXTURE; COMPLEX; THREELEVEL; BASIC; RANDOM; COMPLEX; CROSSCLASSIFIED; RANDOM; EFA # #; BASIC; MIXTURE; COMPLEX; TWOLEVEL; EFA # # UW* # # UB*; EFA # # UW # # UB;	GENERAL

CHAPTER 16

ESTIMATOR =	ML; MLM; MLMV; MLR; MLF; MUML; WLS; WLSM; WLSMV; ULS; ULSMV; GLS; BAYES;	depends on analysis type
MODEL =	CONFIGURAL ; METRIC; SCALAR; NOMEANSTRUCTURE ; NOCOVARIANCES ; ALLFREE ;	means covariances equal
ALIGNMENT =	FIXED; FIXED (reference class CONFIGURAL); FIXED (reference class BSEM); FREE;	last class CONFIGURAL last class CONFIGURAL
DISTRIBUTION =	FREE (reference class CONFIGURAL); FREE (reference class BSEM); NORMAL ; SKEWNORMAL ; TDISTRIBUTION ; SKEWT;	NORMAL
PARAMETERIZATION =	DELTA; THETA; LOGIT; LOGLINEAR ; PROBABILITY ; RESCOVARIANCES ;	DELTA LOGIT
LINK =	LOGIT ; PROBIT ;	RESCOV LOGIT

ROTATION =	GEOMIN;	GEOMIN (OBLIQUE value)
	GEOMIN (OBLIQUE value);	
	GEOMIN (ORTHOGONAL value);	
	QUARTIMIN;	OBLIQUE
	CF-VARIMAX;	OBLIQUE
	CF-VARIMAX (OBLIQUE);	
	CF-VARIMAX (ORTHOGONAL);	
	CF-QUARTIMAX;	OBLIQUE
	CF-QUARTIMAX (OBLIQUE);	
	CF-QUARTIMAX (ORTHOGONAL);	
	CF-EQUAMAX;	OBLIQUE
	CF-EQUAMAX (OBLIQUE);	
	CF-EQUAMAX (ORTHOGONAL);	
	CF-PARSIMAX;	OBLIQUE
	CF-PARSIMAX (OBLIQUE);	
	CF-PARSIMAX (ORTHOGONAL);	
	CF-FACPARSIM;	OBLIQUE
	CF-FACPARSIM (OBLIQUE);	
	CF-FACPARSIM (ORTHOGONAL);	
	CRAWFER;	OBLIQUE 1/p
	CRAWFER (OBLIQUE value);	
	CRAWFER (ORTHOGONAL value);	
	OBLIMIN;	OBLIQUE 0
	OBLIMIN (OBLIQUE value);	
	OBLIMIN (ORTHOGONAL value);	
	VARIMAX;	
	PROMAX;	
	TARGET;	
	BI-GEOMIN;	OBLIQUE
	BI-GEOMIN (OBLIQUE);	
	BI-GEOMIN (ORTHOGONAL);	
	BI-CF-QUARTIMAX;	OBLIQUE
	BI-CF-QUARTIMAX (OBLIQUE);	
	BI-CF-QUARTIMAX (ORTHOGONAL);	
ROWSTANDARDIZATION =	CORRELATION;	CORRELATION
	KAISER;	
	COVARIANCE;	
PARALLEL =	number;	0
REPSE =	BOOTSTRAP;	
	JACKKNIFE;	
	JACKKNIFE1;	
	JACKKNIFE2;	
	BRR;	
	FAY (#);	.3

CHAPTER 16

BASEHAZARD =	ON; OFF; ON (EQUAL); ON (UNEQUAL); OFF (EQUAL); OFF (UNEQUAL);	depends on analysis type EQUAL EQUAL
CHOLESKY =	ON; OFF;	depends on analysis type
ALGORITHM =	EM; EMA; FS; ODLL; INTEGRATION ;	depends on analysis type
INTEGRATION =	number of integration points; STANDARD (number of integration points) ; GAUSSHERMITE (number of integration points) ; MONTECARLO (number of integration points);	STANDARD depends on analysis type 15 depends on analysis type
MCSEED =	random seed for Monte Carlo integration;	0
ADAPTIVE =	ON; OFF;	ON
INFORMATION =	OBSERVED ; EXPECTED ; COMBINATION ;	depends on analysis type
BOOTSTRAP =	number of bootstrap draws; number of bootstrap draws (STANDARD); number of bootstrap draws (RESIDUAL);	STANDARD
LRTBOOTSTRAP =	number of bootstrap draws for TECH14;	depends on analysis type
STARTS =	number of initial stage starts and number of final stage optimizations;	depends on analysis type
STITERATIONS =	number of initial stage iterations;	10
STCONVERGENCE =	initial stage convergence criterion;	1
STSCALE =	random start scale;	5
STSEED =	random seed for generating random starts;	0
OPTSEED =	random seed for analysis;	
K-1STARTS =	number of initial stage starts and number of final stage optimizations for the k-1 class model for TECH14;	20 4
LRTSTARTS =	number of initial stage starts and number of final stage optimizations for TECH14;	0 0 40 8
RSTARTS =	number of random starts for the rotation algorithm and number of factor solutions printed for exploratory factor analysis;	depends on analysis type

ANALYSIS Command

ASTARTS =	number of random starts for the alignment optimization;	30
H1STARTS =	Number of initial stage starts and number of final stage optimizations for the H1 model;	0 0
DIFFTEST =	file name;	
MULTIPLIER =	file name;	
COVERAGE =	minimum covariance coverage with missing data;	.10
ADDFREQUENCY =	value divided by sample size to add to cells with zero frequency;	.5
ITERATIONS =	maximum number of iterations for the Quasi-Newton algorithm for continuous outcomes;	1000
SDITERATIONS =	maximum number of steepest descent iterations for the Quasi-Newton algorithm for continuous outcomes;	20
H1ITERATIONS =	maximum number of iterations for unrestricted model with missing data;	2000
MITERATIONS =	number of iterations for the EM algorithm;	500
MCITERATIONS =	number of iterations for the M step of the EM algorithm for categorical latent variables;	1
MUITERATIONS =	number of iterations for the M step of the EM algorithm for censored, categorical, and count outcomes;	1
RITERATIONS =	maximum number of iterations in the rotation algorithm for exploratory factor analysis;	10000
AITERATIONS =	maximum number of iterations in the alignment optimization;	5000
CONVERGENCE =	convergence criterion for the Quasi-Newton algorithm for continuous outcomes;	depends on analysis type
H1CONVERGENCE =	convergence criterion for unrestricted model with missing data;	.0001
LOGCRITERION =	likelihood convergence criterion for the EM algorithm;	depends on analysis type
RLOGCRITERION =	relative likelihood convergence criterion for the EM algorithm;	depends on analysis type
MCONVERGENCE =	convergence criterion for the EM algorithm;	depends on analysis type
MCCONVERGENCE =	convergence criterion for the M step of the EM algorithm for categorical latent variables;	.000001
MUCONVERGENCE =	convergence criterion for the M step of the EM algorithm for censored, categorical, and count outcomes;	.000001
RCONVERGENCE =	convergence criterion for the rotation algorithm for exploratory factor analysis;	.00001
ACONVERGENCE =	convergence criterion for the derivatives of the alignment optimization;.	.001

CHAPTER 16

MIXC =	ITERATIONS; CONVERGENCE; M step iteration termination based on number of iterations or convergence for categorical latent variables;	ITERATIONS
MIXU =	ITERATIONS; CONVERGENCE; M step iteration termination based on number of iterations or convergence for censored, categorical, and count outcomes;	ITERATIONS
LOGHIGH =	max value for logit thresholds;	+15
LOGLOW =	min value for logit thresholds;	- 15
UCELLSIZE =	minimum expected cell size;	.01
VARIANCE =	minimum variance value;	.0001
SIMPLICITY =	SQRT; FOURTHRT;	SQRT
TOLERANCE =	simplicity tolerance value;	.0001
METRIC=	REFGROUP; PRODUCT;	REFGROUP
MATRIX =	COVARIANCE; CORRELATION;	COVARIANCE
POINT =	MEDIAN; MEAN; MODE;	MEDIAN
CHAINS =	number of MCMC chains;	2
BSEED =	seed for MCMC random number generation;	0
STVALUES =	UNPERTURBED; PERTURBED; ML;	UNPERTURBED
PREDICTOR =	LATENT; OBSERVED;	LATENT
ALGORITHM =	GIBBS; GIBBS (PX1); GIBBS (PX2); GIBBS (PX3); GIBBS (RW); MH;	GIBBS (PX1)
BCONVERGENCE =	MCMC convergence criterion using Gelman-Rubin PSR;	.05
BITERATIONS =	maximum and minimum number of iterations for each MCMC chain when Gelman-Rubin PSR is used;	50000 0
FBITERATIONS =	fixed number of iterations for each MCMC chain when Gelman-Rubin PSR is not used;	
THIN =	k where every k-th MCMC iteration is saved;	1

MDITERATIONS =	maximum number of iterations used to compute the Bayes multivariate mode;	10000
KOLMOGOROV =	number of draws from the MCMC chains;	100
PRIOR =	number of draws from the prior distribution;	1000
INTERACTIVE =	file name;	
PROCESSORS =	# of processors # of threads;	1 1

The ANALYSIS command is not a required command. Default settings are shown in the last column. If the default settings are appropriate for the analysis, it is not necessary to specify the ANALYSIS command.

Note that commands and options can be shortened to four or more letters. Option settings can be referred to by either the complete word or the part of the word shown above in bold type.

TYPE

The TYPE option is used to describe the type of analysis. There are six major analysis types in Mplus: GENERAL, MIXTURE, TWOLEVEL, THREELEVEL, CROSSCLASSIFIED, and EFA. GENERAL is the default.

The default is to estimate the model under missing data theory using all available data; to include means, thresholds, and intercepts in the model; to compute standard errors; and to compute chi-square when available. These defaults can be overridden. The LISTWISE option of the DATA command can be used to delete all observations from the analysis that have one or more missing values on the set of analysis variables. For TYPE=GENERAL, means, thresholds, and intercepts can be excluded from the analysis model by specifying MODEL=NOMEANSTRUCTURE in the ANALYSIS command. The NOSERROR option of the OUTPUT command can be used to suppress the computation of standard errors. The NOCHISQUARE option of the OUTPUT command can be used to suppress the computation of chi-square. In some models, suppressing the computation of standard errors and chi-square can greatly reduce computational time. Following is a description of each of the four major analysis types.

GENERAL

Analyses using TYPE=GENERAL include models with relationships among observed variables, among continuous latent variables, and among observed variables and continuous latent variables. In these models, the continuous latent variables represent factors and random effects. Observed outcome variables can be continuous, censored, binary, ordered categorical (ordinal), counts, or combinations of these variable types. In addition, for regression analysis and path analysis for non-mediating outcomes, observed outcome variables can be unordered categorical (nominal). Following are models that can be estimated using TYPE=GENERAL:

- Regression analysis
- Path analysis
- Confirmatory factor analysis
- Structural equation modeling
- Growth modeling
- Discrete-time survival analysis
- Continuous-time survival analysis
- N=1 time series analysis

Special features available with the above models for all observed outcome variable types are:

- Multiple group analysis
- 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
- Bootstrap standard errors and confidence intervals
- Wald chi-square test of parameter equalities

Following is a list of the other TYPE settings that can be used in conjunction with TYPE=GENERAL along with a brief description of their functions:

- BASIC computes sample statistics and other descriptive information.
- RANDOM allows models with both random intercepts and random slopes.
- COMPLEX computes standard errors and a chi-square test of model fit taking into account stratification, non-independence of observations, and/or unequal probability of selection.

Following is an example of how to specify the TYPE option for a regression analysis with a random slope:

```
TYPE = GENERAL RANDOM;
```

or simply,

```
TYPE = RANDOM;
```

because GENERAL is the default.

MIXTURE

Analyses using TYPE=MIXTURE include models with categorical latent variables which are also referred to as latent class or finite mixture models.

For models with only categorical latent variables, observed outcome variables can be continuous, censored, binary, ordered categorical (ordinal), unordered categorical (nominal), counts, or combinations of these variable types. Following are the models that can be estimated using TYPE=MIXTURE with only categorical latent variables:

- Regression mixture modeling
- Path analysis mixture modeling
- Latent class analysis
- Latent class analysis with covariates and direct effects
- Confirmatory latent class analysis
- Latent class analysis with multiple categorical latent variables
- Loglinear modeling

CHAPTER 16

- Non-parametric modeling of latent variable distributions
- Finite mixture modeling
- Complier Average Causal Effect (CACE) modeling
- Latent transition analysis and hidden Markov modeling including mixtures and covariates
- Latent class growth analysis
- Discrete-time survival mixture analysis
- Continuous-time survival mixture analysis

For models that include both continuous and categorical latent variables, observed outcome variables can be continuous, censored, binary, ordered categorical (ordinal), counts, or combinations of these variable types. In addition, for regression analysis and path analysis for non-mediating outcomes, observed outcome variables can also be unordered categorical (nominal). Following are models that can be estimated using TYPE=MIXTURE with both continuous and categorical latent variables:

- Latent class analysis with random effects
- Factor mixture modeling
- SEM mixture modeling
- Growth mixture modeling with latent trajectory classes
- Discrete-time survival mixture analysis
- Continuous-time survival mixture analysis

Special features available with the above models for all observed outcome variable types are:

- Multiple group analysis
- 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
- Bootstrap standard errors and confidence intervals
- Wald chi-square test of parameter equalities
- Analysis with between-level categorical latent variables

- Test of equality of means across latent classes using posterior probability-based multiple imputations

Following is a list of the other TYPE settings that can be used in conjunction with TYPE=MIXTURE along with a brief description of their functions:

- BASIC computes sample statistics and other descriptive information.
- RANDOM allows models with both random intercepts and random slopes.
- COMPLEX computes standard errors and a chi-square test of model fit taking into account stratification, non-independence of observations, and/or unequal probability of selection.

TWOLEVEL

Analyses using TYPE=TWOLEVEL include models with random intercepts and random slopes that vary across clusters in hierarchical data. These random effects can be specified for any of the relationships of the multilevel modeling framework. Observed outcome variables can be continuous, censored, binary, ordered categorical (ordinal), unordered categorical (nominal), counts, or combinations of these variable types.

Special features available for two-level models for all observed outcome variable types are:

- Multiple group analysis
- 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

Following is a list of the other TYPE settings that can be used in conjunction with TYPE=TWOLEVEL along with a brief description of their functions:

- BASIC computes sample statistics and other descriptive information.
- RANDOM allows models with random intercepts, random slopes, random factor loadings, and random variances.
- MIXTURE allows models that have both categorical and continuous latent variables.
- COMPLEX computes standard errors and a chi-square test of model fit taking into account stratification, non-independence of observations, and/or unequal probability of selection.

THREELEVEL

Analyses using TYPE=THREELEVEL include models with random intercepts and random slopes that vary across clusters in hierarchical data. These random effects can be specified for any of the relationships of the multilevel modeling framework. Observed outcome variables can be continuous, binary, or combinations of these variable types.

Special features available for three-level models for all observed outcome variable types are:

- Multiple group analysis
- Missing data
- Complex survey data
- Random slopes
- Linear and non-linear parameter constraints
- Maximum likelihood estimation for continuous outcomes
- Wald chi-square test of parameter equalities

Following is a list of the other TYPE settings that can be used in conjunction with TYPE=THREELEVEL along with a brief description of their functions:

- BASIC computes sample statistics and other descriptive information.
- RANDOM allows models with random intercepts, random slopes, random factor loadings, and random variances.

- **COMPLEX** computes standard errors and a chi-square test of model fit taking into account stratification, non-independence of observations, and/or unequal probability of selection. This is available for continuous variables.

CROSSCLASSIFIED

Analyses using **TYPE=CROSSCLASSIFIED** include models with random intercepts, random slopes, and random variances that vary across clusters in hierarchical data. These random effects can be specified for any of the relationships of the multilevel modeling framework. Observed outcome variables can be continuous, binary, ordered categorical (ordinal), or combinations of these variable types.

Special features available for cross-classified models for all observed outcome variable types are:

- Missing data
- Random slopes
- Random variances

Following is a list of the other **TYPE** settings that can be used in conjunction with **TYPE=CROSSCLASSIFIED** along with a brief description of their functions:

- **RANDOM** allows models with random intercepts, random slopes, and random variances.

EFA

Analyses using **TYPE=EFA** include exploratory factor analysis of continuous, censored, binary, ordered categorical (ordinal), counts, or combinations of these variable types. See the **ROTATION** option of the **ANALYSIS** command for a description of the rotations available for **TYPE=EFA**. Modification indices are available for the residual correlations using the **MODINDICES** option of the **OUTPUT** command.

Special features available for EFA for all observed outcome variable types are:

- Missing data

CHAPTER 16

- Complex survey data

Following are the other TYPE settings that can be used in conjunction with TYPE=EFA along with a brief description of their functions:

- BASIC computes sample statistics and other descriptive information.
- MIXTURE allows models that have both categorical and continuous latent variables.
- COMPLEX computes standard errors and a chi-square test of model fit taking into account stratification, non-independence of observations, and/or unequal probability of selection.
- TWOLEVEL models non-independence of observations due to clustering taking into account stratification and/or unequal probability of selection.

Following is an example of how to specify the TYPE option for a single-level exploratory factor analysis:

```
TYPE = EFA 1 3;
```

where the two numbers following EFA are the lower and upper limits of the number of factors to be extracted. In the example above factor solutions are given for one, two, and three factors.

Following is an example of how to specify the full TYPE option for a multilevel exploratory factor analysis (Asparouhov & Muthén, 2007):

```
TYPE = TWOLEVEL EFA 3 4 UW* 1 2 UB*;
```

where the first two numbers, 3 and 4, are the lower and upper limits of the number of factors to be extracted on the within level, UW* specifies that an unrestricted within-level model is estimated, the second two numbers, 1 and 2, are the lower and upper limits of the number of factors to be extracted on the between level, and UB* specifies that an unrestricted between-level model is estimated. The within- and between-level specifications are crossed. In the example shown above, the three- and four-factor models and the unrestricted model on the within level are estimated in combination with the one- and two-factor models and the unrestricted model on between resulting in nine solutions.

If `UW` and `UB` are used instead of `UW*` and `UB*`, the unrestricted models are not estimated but instead the model parameters are fixed at the sample statistic values. This can speed up the analysis.

For multilevel exploratory factor analysis, the `TYPE` option can be specified using only numbers or the `UW*` and `UB*` specifications for each level. For example,

```
TYPE = TWOLEVEL EFA 3 4 UB*;
```

specifies that three- and four-factors models on the within level are estimated in combination with an unrestricted model on the between level.

```
TYPE = TWOLEVEL EFA UW* 12;
```

specifies that an unrestricted model on the within level is estimated in combination with one- and two-factor model on the between level.

ESTIMATOR

The `ESTIMATOR` option is used to specify the estimator to be used in the analysis. The default estimator differs depending on the type of analysis and the measurement scale of the dependent variable(s). Not all estimators are available for all models. Following is a table that shows which estimators are available for specific models and variable types. The information is broken down by models with all continuous dependent variables, those with at least one binary or ordered categorical dependent variable, and those with at least one censored, unordered categorical, or count dependent variable. All of the estimators require individual-level data except `ML` for `TYPE=GENERAL` and `EFA`, `GLS`, and `ULS` which can use summary data. The default settings are indicated by bold type.

The first column of the table shows the combinations of `TYPE` settings that are allowed. The second column shows the set of estimators available for the analysis types in the first column for a model with all continuous dependent variables. The third column shows the set of estimators available for the analysis types in the first column for a model with at least one binary or ordered categorical dependent variable. The fourth column shows the set of estimators available for the analysis

CHAPTER 16

types in the first column for a model with at least one censored, unordered categorical, or count dependent variable.

Type of Analysis TYPE=	All continuous dependent variables	At least one binary or ordered categorical dependent variable	At least one censored, unordered categorical, or count dependent variable
GENERAL	ML** MLM**** MLMV**** MLR** MLF** GLS**** WLS**** BAYES	WLS WLSM WLSMV ULSMV ML* MLR* MLF* BAYES	WLS**** WLSM**** WLSMV**** ML* MLR* MLF*
GENERAL RANDOM	ML** MLR** MLF**	ML* MLR* MLF*	ML* MLR* MLF*
GENERAL RANDOM COMPLEX	MLR**	MLR*	MLR*
GENERAL COMPLEX	ML**** MLR**	WLS WLSM WLSMV ULSMV MLR*	WLS**** WLSM**** WLSMV**** ULSMV**** MLR*
MIXTURE	ML** MLR** MLF** BAYES	ML** MLR** MLF** BAYES	ML** MLR** MLF**
MIXTURE RANDOM	ML** MLR** MLF**	ML** MLR** MLF**	ML** MLR** MLF**
MIXTURE COMPLEX MIXTURE COMPLEX RANDOM	MLR**	MLR**	MLR*
TWOLEVEL	MUML*** ML** MLR** MLF** WLS WLSM WLSMV ULSMV BAYES	ML* MLR* MLF* WLS WLSM WLSMV ULSMV BAYES	ML* MLR* MLF*
TWOLEVEL RANDOM	ML** MLR** MLF** BAYES	ML* MLR* MLF* BAYES	ML* MLR* MLF*
TWOLEVEL MIXTURE TWOLEVEL RANDOM MIXTURE	ML* MLR* MLF*	ML* MLR* MLF*	ML* MLR* MLF*
COMPLEX TWOLEVEL COMPLEX TWOLEVEL RANDOM	MLR**	MLR*	MLR*

ANALYSIS Command

COMPLEX TWOLEVEL MIXTURE COMPLEX TWOLEVEL RANDOM MIXTURE	MLR*	MLR*	MLR*
THREELEVEL THREELEVEL RANDOM	ML MLR MLF BAYES	BAYES	NA
COMPLEX THREELEVEL COMPLEX THREELEVEL RANDOM	MLR	NA	NA
CROSSCLASSIFIED CROSSCLASSIFIED RANDOM	BAYES	BAYES	NA
EFA	ML MLR** MLF** ULS***** BAYES	WLS WLSM WLSMV ULS ULSMV ML* MLR* MLF* BAYES	ML* MLR* MLF*
EFA MIXTURE	ML** MLR** MLF**	ML* MLR* MLF*	ML* MLR* MLF*
EFA COMPLEX	MLR**	WLS WLSM WLSMV ULSMV MLR*	MLR*
EFA TWOLEVEL	M** MLR** MLF** WLS WLSM WLSMV ULSMV	WLS WLSM WLSMV ULSMV	NA

- * Numerical integration required
- ** Numerical integration an option
- *** Maximum likelihood with balanced data, limited-information for unbalanced data, not available with missing data
- **** Only available for censored outcomes without inflation
- ***** Not available with missing data
- ***** Default with replicate weights
- NA Not available

Following is a description of what the above estimator settings represent:

- ML – maximum likelihood parameter estimates with conventional standard errors and chi-square test statistic
- MLM – maximum likelihood parameter estimates with standard errors and a mean-adjusted chi-square test statistic that are robust to

non-normality. The MLM chi-square test statistic is also referred to as the Satorra-Bentler chi-square.

- MLMV – maximum likelihood parameter estimates with standard errors and a mean- and variance-adjusted chi-square test statistic that are robust to non-normality
- MLR – maximum likelihood parameter estimates with standard errors and a chi-square test statistic (when applicable) that are robust to non-normality and non-independence of observations when used with TYPE=COMPLEX. The MLR standard errors are computed using a sandwich estimator. The MLR chi-square test statistic is asymptotically equivalent to the Yuan-Bentler T2* test statistic.
- MLF – maximum likelihood parameter estimates with standard errors approximated by first-order derivatives and a conventional chi-square test statistic
- MUML – Muthén’s limited information parameter estimates, standard errors, and chi-square test statistic
- WLS – weighted least square parameter estimates with conventional standard errors and chi-square test statistic that use a full weight matrix. The WLS chi-square test statistic is also referred to as ADF when all outcome variables are continuous.
- WLSM – weighted least square parameter estimates using a diagonal weight matrix with standard errors and mean-adjusted chi-square test statistic that use a full weight matrix
- WLSMV – weighted least square parameter estimates using a diagonal weight matrix with standard errors and mean- and variance-adjusted chi-square test statistic that use a full weight matrix
- ULS – unweighted least squares parameter estimates
- ULSMV – unweighted least squares parameter estimates with standard errors and a mean- and variance-adjusted chi-square test statistic that use a full weight matrix
- GLS – generalized least square parameter estimates with conventional standard errors and chi-square test statistic that use a normal-theory based weight matrix
- Bayes – Bayesian posterior parameter estimates with credibility intervals and posterior predictive checking

BAYESIAN ESTIMATION

Bayesian estimation differs from frequentist estimation in that parameters are not considered to be constants but to be variables

(Gelman et al., 2004). The parameters can be given priors corresponding to theory or previous studies. Together with the likelihood of the data, this gives rise to posterior distributions for the parameters. Bayesian estimation uses Markov chain Monte Carlo (MCMC) algorithms to create approximations to the posterior distributions by iteratively making random draws in the MCMC chain. The initial draws in the MCMC chain are referred to as the burnin phase. In Mplus, the first half of each chain is discarded as being part of the burnin phase. Convergence is assessed using the Gelman-Rubin convergence criterion based on the potential scale reduction factor for each parameter (Gelman & Rubin, 1992; Gelman et al., 2004, pp. 296-297). With multiple chains, this is a comparison of within- and between-chain variation. With a single chain, the last half of the iterations is split into two quarters and the potential scale reduction factor is computed for these two quarters. Convergence can also be monitored by the trace plots of the posterior draws in the chains. Auto-correlation plots describe the degree of non-independence of consecutive draws. These plots aid in determining the quality of the mixing in the chain. For each parameter, credibility intervals are obtained from the percentiles of its posterior distribution. Model comparisons are aided by the Deviance Information Criterion (DIC). Overall test of model fit is judged by Posterior Predictive Checks (PPC) where the observed data is compared to the posterior predictive distribution. In Mplus, PPC p-values are computed using the likelihood-ratio chi-square statistic for continuous outcomes and for the continuous latent response variables of categorical outcomes. Gelman et al. (2004, Chapter 6) and Lee (2007, Chapter 5) give overviews of model comparison and model checking. For a technical description of the Bayesian implementation, see Asparouhov and Muthén (2010b). See also Chapter 9 of Muthén, Muthén, and Asparouhov (2016).

Bayesian estimation is available for continuous, binary, ordered categorical (ordinal) or combinations of these variable types with `TYPE=GENERAL`, `MIXTURE` with only one categorical latent variable, `TWOLEVEL`, `THREELEVEL`, `THREELEVEL RANDOM`, `CROSSCLASSIFIED`, `CROSSCLASSIFIED RANDOM`, and `EFA`. To obtain Bayesian estimation, specify:

```
ESTIMATOR=BAYES;
```

MODEL

The MODEL option has two uses. The first use is to automatically set up multiple group models for the purpose of testing for measurement invariance. The second use is to make changes to the defaults of the MODEL command.

TESTING FOR MEASUREMENT INVARIANCE

The MODEL option is used to automatically set up multiple group models for the purpose of testing for measurement invariance using the GROUPING option or the KNOWNCLASS option. It is available for CFA and ESEM models for continuous variables with the maximum likelihood and Bayes estimators; for censored variables with the weighted least squares and maximum likelihood estimators; for binary and ordered categorical (ordinal) variables using the weighted least squares, maximum likelihood, and Bayes estimators; and for count variables using the maximum likelihood estimator. It is not available for censored-inflated, count-inflated, nominal, continuous-time survival, negative binomial variables, or combinations of variable types. The MODEL command can contain only BY statements for first-order factors. The metric for the factors can be set by fixing a factor loading to one in each group or by fixing the factor variance to one in one group. No partial measurement invariance is allowed. The configural, metric, and scalar models used are described in Chapter 14.

The MODEL option has three settings for testing for measurement invariance: CONFIGURAL, METRIC, and SCALAR. These settings can be used alone to set up a particular model or together to test the models for measurement invariance. Chi-square difference testing is carried out automatically using scaling correction factors for MLM, MLR, and WLSM and using the DIFFTEST option for WLSMV and MLMV. The settings cannot be used together for ESTIMATOR=BAYES and for Monte Carlo analyses. Full analysis results are printed along with a summary of the difference testing. The CONFIGURAL setting produces a model with the same number of factors and the same set of zero factor loadings in all groups. The METRIC setting produces a model where factor loadings are held equal across groups. The SCALAR setting produces a model where factor loadings and intercepts/thresholds are held equal across groups. When the factor variance is fixed to one in one group, it is the first group when

the `GROUPING` option is used and the last class when the `KNOWNCLASS` option is used.

The `MODEL` option for testing measurement invariance is specified as follows:

```
MODEL = CONFIGURAL METRIC SCALAR;
```

which specifies that configural, metric, and scalar models will be estimated and difference testing of the models will be done.

CHANGE DEFAULTS OF MODEL COMMAND

The `MODEL` option has three settings that change the defaults of the `MODEL` command: `NOMEANSTRUCTURE`, `NOCOVARIANCES`, and `ALLFREE`. The `NOMEANSTRUCTURE` setting is used with `TYPE=GENERAL` to specify that means, intercepts, and thresholds are not included in the analysis model. The `NOCOVARIANCES` setting specifies that the covariances and residual covariances among all latent and observed variables in the analysis model are fixed at zero. The `WITH` option of the `MODEL` command can be used to free selected covariances and residual covariances. Following is an example of how to specify that the covariances and residual covariances among all latent and observed variables in the model are fixed at zero:

```
MODEL = NOCOVARIANCES;
```

The `ALLFREE` setting is used with `TYPE=MIXTURE`, the `KNOWNCLASS` option, `ESTIMATOR=BAYES`, and a special automatic labeling function to assign zero-mean and small-variance priors to differences in intercepts, thresholds, and factor loadings across groups. By specifying `MODEL=ALLFREE`, factor means, variances, and covariances are free across groups except for factor means in the last group which are fixed at zero. In addition, intercepts, thresholds, factor loadings, and residual variances of the factor indicators are free across the groups.

Following is an example of how to use the `ALLFREE` setting and automatic labeling to assign zero-mean and small-variance priors to differences in intercepts, thresholds, and factor loadings across the ten groups.

```

MODEL:      %OVERALL%
            f1 BY y1-y3* (lam#_1-lam#_3);
            f2 BY y4-y6* (lam#_4-lam#_6);
            [y1-y6] (nu#_1-nu#_6);

```

```

MODEL PRIORS:
            DO(1,6) DIFF(lam1_#-lam10_#)~N(0,0.01);
            DO(1,6) DIFF(nu1_#-nu10_#)~N(0,0.01);

```

In the overall part of the model, labels are assigned to the factor loadings and the intercepts using automatic labeling for groups. The labels must include the number sign (#) followed by the underscore (_) symbol followed by a number. The number sign (#) refers to a group and the number refers to a parameter. The label lam#_1 is assigned to the factor loading for y1; the label lam#_2 is assigned to the factor loading for y2; and the label lam#_3 is assigned to the factor loading for y3. These labels are expanded to include group information. For example, the label for parameter 1 is expanded across the ten groups to give labels lam1_1, lam2_1 through lam10_1. In MODEL PRIORS, these expanded labels are used to assign zero-mean and small-variance priors to the differences across groups of the factor loadings and intercepts using the DO and DIFF options. They can be used together to simplify the assignment of priors to a large set of difference parameters for models with multiple groups and multiple time points. For the DO option, the numbers in parentheses give the range of values for the do loop. The number sign (#) is replaced by these values during the execution of the do loop. The numbers refer to the six factor indicators.

ALIGNMENT

The ALIGNMENT option is used with multiple group models to assess measurement invariance and compare factor means and variances across groups (Asparouhov & Muthén, 2014c). It is most useful when there are many groups as seen in country comparisons of achievement like the Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS) as well as in cross-cultural studies like the International Social Survey Program (ISSP) and the European Social Survey (ESS). It is available when all variables are continuous or binary with the ML, MLR, MLF, and BAYES estimators and when all variables are ordered categorical (ordinal) with the ML,

MLR, and MLF estimators. It is available for regular and Monte Carlo analyses using TYPE=MIXTURE and TYPE=COMPLEX MIXTURE in conjunction with the KNOWNCLASS option for real data and the NGROUPS option for Monte Carlo analyses. The MODEL command can contain only BY statements for first-order factors where factor indicators do not load on more than one factor.

The alignment optimization method consists of three steps:

1. Analysis of a configural model with the same number of factors and same pattern of zero factor loadings in all groups.
2. Alignment optimization of the measurement parameters, factor loadings and intercepts/thresholds according to a simplicity criterion that favors few non-invariant measurement parameters.
3. Adjustment of the factor means and variances in line with the optimal alignment.

The ALIGNMENT option has two settings: FIXED and FREE. There is no default. In the FIXED setting, a factor mean is fixed at zero in the reference group. In the FREE setting, all factor means are estimated. FREE is the most general approach. FIXED is recommended when there is little factor loading non-invariance which may occur when there is a small number of groups. The ALIGNMENT option has two subsettings for specifying the reference group and the type of configural model used in the first step of the alignment optimization. The default for the reference group is the last known class. The default for the type of configural model is CONFIGURAL. The alternative setting is BSEM where approximate invariance of measurement parameters is specified using Bayes priors (Muthén & Asparouhov, 2013). The subsettings are specified in parentheses following the FIXED or FREE settings. Following is an example of how to specify the ALIGNMENT option:

```
ALIGNMENT = FREE;
```

where the default reference group is the last known class. The default configural model is CONFIGURAL. Following are three equivalent ways to specify this:

```
ALIGNMENT = FREE (1 CONFIGURAL);
```

```
ALIGNMENT = FREE (1);
```

ALIGNMENT = FREE (CONFIGURAL);

DISTRIBUTION

The **DISTRIBUTION** option is used in conjunction with **TYPE=GENERAL** and **TYPE=MIXTURE** to specify non-normal distributions for continuous observed variables and continuous factors (Asparouhov & Muthén, 2015a; Muthén & Asparouhov, 2015a). These new methods are experimental in that they have not been extensively used in practice.

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

The **SKEWNORMAL** and **SKEWT** settings have a special skew parameter for each observed and latent variable that is related to the skewness of the variable. It is specified by mentioning the name of the variable in curly brackets. For example, the skew parameter for a variable *y* is specified as {*y*}. Skew parameters are free and unequal across groups or classes with starting values of one. They can be constrained to be equal or fixed at a particular value. The **TDISTRIBUTION** and **SKEWT** settings have a special degree of freedom parameter for each group or class that is related to the degrees of freedom in the t-distribution. The degree of freedom parameter is specified by putting *df* in curly brackets, for example, {*df*}. The degree of freedom parameter is free and unequal across groups or classes with a starting value of one. It can be constrained to be equal or fixed at a particular value. The **SKEWNORMAL** setting can capture skewness less than the absolute value of one, whereas the **SKEWT** setting has no such limitations.

PARAMETERIZATION

The **PARAMETERIZATION** option is used for three purposes. The first purpose is to change from the default Delta parameterization to the alternative Theta parameterization when **TYPE=GENERAL** is used, at

least one observed dependent variable is categorical, and weighted least squares estimation is used in the analysis. The second purpose is to change from the default logit regression parameterization to either the loglinear or probability parameterization when TYPE=MIXTURE and more than one categorical latent variable is used in the analysis. The third purpose is to allow the WITH option of the MODEL command to be used to specify residual covariances for binary and ordered categorical (ordinal) outcomes using maximum likelihood estimation.

DELTA VERSUS THETA PARAMETERIZATION

There are two model parameterizations available when TYPE=GENERAL is used, one or more dependent variables are categorical, and weighted least squares estimation is used in the analysis. The first parameterization is referred to as DELTA. This is the default parameterization. In the DELTA parameterization, scale factors for continuous latent response variables of observed categorical outcome variables are allowed to be parameters in the model, but residual variances for continuous latent response variables are not. The second parameterization is referred to as THETA. In the THETA parameterization, residual variances for continuous latent response variables of observed categorical outcome variables are allowed to be parameters in the model, but scale factors for continuous latent response variables are not.

The DELTA parameterization is the default because it has been found to perform better in many situations (Muthén & Asparouhov, 2002). The THETA parameterization is preferred when hypotheses involving residual variances are of interest. Such hypotheses may arise with multiple group analysis and analysis of longitudinal data. In addition, there are certain models that can be estimated using only the THETA parameterization because they have been found to impose improper parameter constraints with the DELTA parameterization. These are models where a categorical dependent variable is both influenced by and influences either another observed dependent variable or a latent variable.

To select the THETA parameterization, specify the following:

```
PARAMETERIZATION = THETA;
```

LOGIT, LOGLINEAR, AND PROBABILITY PARAMETERIZATIONS

There are three model parameterizations available when TYPE=MIXTURE is used and more than one categorical latent variable is used in the analysis. The first parameterization is referred to as LOGIT. This is the default parameterization. In the LOGIT parameterization, logistic regressions are estimated for categorical latent variables. In the LOGIT parameterization, the ON and WITH options of the MODEL command can be used to specify the relationships between the categorical latent variables. The second parameterization is referred to as LOGLINEAR. In the LOGLINEAR parameterization, loglinear models are estimated for categorical latent variables allowing two- and three-way interactions. In the LOGLINEAR parameterization, only the WITH option of the MODEL command can be used to specify the relationships between the categorical latent variables. The third parameterization is referred to as PROBABILITY. In the PROBABILITY parameterization, categorical latent variable regression coefficients can be expressed as probabilities rather than logits.

RESCOVARIANCES

The RESCOVARIANCES option is used with Latent Class Analysis and Latent Transition Analysis to specify residual covariances for binary and ordered categorical (ordinal) variables using maximum likelihood estimation (Asparouhov & Muthén, 2015b). These residual covariances are specified using the WITH option of the MODEL command. They can be free across classes, constrained to be equal across classes, or appear in only certain classes. Following is a partial input for a latent class analysis where the residual covariances are held equal across classes:

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

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

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

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

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

LINK

The LINK option is used with maximum likelihood estimation to select a logit or probit link for models with categorical outcomes. The default is a logit link. Following is an example of how to request a probit link:

```
LINK = PROBIT;
```

ROTATION

The ROTATION option is used with TYPE=EFA to specify the type of rotation of the factor loading matrix to be used in exploratory factor analysis. The default is the GEOMIN oblique rotation (Yates, 1987; Browne, 2001). The algorithms used in the rotations are described in Jennrich and Sampson (1966), Browne (2001), Bernaards and Jennrich (2005), and Jennrich (2007). For consistency, the names of the rotations used in the CEFA program (Browne, Cudeck, Tateneni, & Mels, 2004) are used for rotations that are included in both the CEFA and Mplus programs. Target rotations (Browne, 2001) and bi-factor rotations (Jennrich & Bentler, 2011, 2012) are also available.

Standard errors are available as the default for all rotations except PROMAX and VARIMAX. THE NOSERROR options of the OUTPUT command can be used to request that standard errors not be computed. The following rotations are available:

GEOMIN
 QUARTIMIN
 CF-VARIMAX
 CF-QUARTIMAX
 CF-EQUAMAX
 CF-PARSIMAX
 CF-FACPARSIM
 CRAWFER
 OBLIMIN
 PROMAX
 VARIMAX
 TARGET
 BI-GEOMIN
 BI-CF-QUARTIMAX

All rotations are available as both oblique and orthogonal except PROMAX and QUARTIMIN which are oblique and VARIMAX which is orthogonal. The default for rotations that can be both oblique and orthogonal is oblique.

The GEOMIN rotation is recommended when factor indicators have substantial loadings on more than one factor resulting in a variable complexity greater than one. Geomin performs well on Thurstone's 26

variable Box Data (Browne, 2001, Table 3, p. 135). The GEOMIN epsilon (Browne, 2001) default setting varies as a function of the number of factors. With two factors, it is .0001. With three factors, it is .001. With four or more factors, it is .01. The default can be overridden using the GEOMIN option. The epsilon value must be a positive number. The Geomin rotation algorithm often finds several local minima of the rotation function (Browne, 2001). To find a global minimum, 30 random rotation starts are used as the default. The RSTARTS option of the ANALYSIS command can be used to change the default.

Following is an example of how to change the GEOMIN epsilon value for an oblique rotation:

```
ROTATION = GEOMIN (OBLIQUE .5);
```

or

```
ROTATION = GEOMIN (.5);
```

where .5 is the value of epsilon.

Following is an example of how to specify an orthogonal rotation for the GEOMIN rotation and to specify an epsilon value different from the default:

```
ROTATION = GEOMIN (ORTHOGONAL .5);
```

The QUARTIMIN rotation uses the direct quartimin rotation of Jennrich and Sampson (1966). The following rotations are identical to direct quartimin:

```
CF-QUARTIMAX (OBLIQUE)
CRAWFER (OBLIQUE 0)
OBLIMIN (OBLIQUE 0)
```

The rotations that begin with CF are part of the Crawford-Ferguson family of rotations (Browne, 2001). They are related to the CRAWFER rotation by the value of the CRAWFER parameter kappa. Following are the values of kappa for the Crawford-Ferguson family of rotations (Browne, 2001, Table 1):

CHAPTER 16

CF-VARIMAX	1/p
CF-QUARTIMAX	0
CF-EQUAMAX	m/2p
CF-PARSIMAX	(m-1)/(p+m-2)
CF-FACPARSIM	1

where p is the number of variables and m is the number of factors.

The default for these rotations is oblique. Following is an example of how to specify an orthogonal rotation for the Crawford-Ferguson family of rotations:

```
ROTATION = CF-VARIMAX (ORTHOGONAL);
```

The CRAWFER rotation is a general form of the Crawford-Ferguson family of rotations where kappa can be specified as a value from 0 through 1. The default value of kappa is 1/p where p is the number of variables. Following is an example of how to specify an orthogonal rotation for the CRAWFER rotation and to specify a kappa value different from 1/p:

```
ROTATION = CRAWFER (ORTHOGONAL .5);
```

where .5 is the value of kappa. The kappa value can also be changed for an oblique rotation as follows:

```
ROTATION = CRAWFER (OBLIQUE .5);
```

or

```
ROTATION = CRAWFER (.5);
```

The default for the OBLIMIN rotation is oblique with a gamma value of 0. Gamma can take on any value. Following is an example of how to specify an orthogonal rotation for the OBLIMIN rotation and to specify a gamma value different than 0:

```
ROTATION = OBLIMIN (ORTHOGONAL 1);
```

where 1 is the value of gamma. The gamma value can also be changed for an oblique rotation as follows:

ROTATION = OBLIMIN (OBLIQUE 1);

or

ROTATION = OBLIMIN (1);

The VARIMAX and PROMAX rotations are the same rotations as those available in earlier versions of Mplus. The VARIMAX rotation is the same as the CF-VARIMAX orthogonal rotation except that VARIMAX row standardizes the factor loading matrix before rotation.

The TARGET setting of the ROTATION option (Browne, 2001) is used with models that have a set of EFA factors in the MODEL command. This setting allows the specification of target factor loading values to guide the rotation of the factor loading matrix. Typically these values are zero. The default for the TARGET rotation is oblique. Following is an example of how to specify an orthogonal TARGET rotation:

ROTATION = TARGET (ORTHOGONAL);

For TARGET rotation, a minimum number of target values must be given for purposes of model identification. For the oblique TARGET rotation, the minimum is $m(m-1)$ where the m is the number of factors. For the orthogonal TARGET rotation, the minimum is $m(m-1)/2$. The target values are given in the MODEL command using the tilde (~) symbol, for example,

```
f1 BY y1-y10 y1~0 (*t);
f2 BY y1-y10 y5~0 (*t);
```

where the target values of y_1 and y_5 are zero.

For the bi-factor rotations, BI-GEOMIN and BI-CF-QUARTIMAX, a general factor is allowed in combination with specific factors. In the oblique rotation, the specific factors are correlated with the general factor and are correlated with each other. In the orthogonal rotation, the specific factors are uncorrelated with the general factor and are uncorrelated with each other. The default for the BI-GEOMIN and BI-CF-QUARTIMAX rotations is oblique. Following is an example of how to specify an orthogonal BI-GEOMIN rotation:

ROTATION = BI-GEOMIN (ORTHOGONAL);

ROWSTANDARDIZATION

The ROWSTANDARDIZATION option is used with exploratory factor analysis (EFA) and when a set of EFA factors is part of the MODEL command to request row standardization of the factor loading matrix before rotation. The ROWSTANDARDIZATION option has three settings: CORRELATION, KAISER, and COVARIANCE. The CORRELATION setting rotates a factor loading matrix derived from a correlation matrix with no row standardization. The KAISER setting rotates a factor loading matrix derived from a correlation matrix with standardization of the factor loadings in each row using the square root of the sum of the squares of the factor loadings in each row (Browne, 2001). The COVARIANCE setting rotates a factor loading matrix derived from a covariance matrix with no row standardization. The default is CORRELATION. The COVARIANCE setting is not allowed for TYPE=EFA. If factor loading equalities are specified in a model for EFA factors, the CORRELATION and KAISER settings are not allowed.

Following is an example of how to specify row standardization using the Kaiser method:

ROWSTANDARDIZATION = KAISER;

PARALLEL

The PARALLEL option is used with TYPE=EFA to determine the optimum number of factors in an exploratory factor analysis. It is available for continuous outcomes using maximum likelihood estimation. Parallel analysis (see, for example, Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hayton, Allen, & Scarpello, 2004) is a method that uses random data with the same number of observations and variables as the original data. The correlation matrix of the random data is used to compute eigenvalues. These eigenvalues are compared to the eigenvalues of the original data. The optimum number of factors is the number of the original data eigenvalues that are larger than the random data eigenvalues. TYPE=PLOT2 of the PLOT command gives a plot of the sample eigenvalues, the parallel analysis eigenvalues, and the

parallel analysis eigenvalues for the 95th percentile. The PARALLEL option is specified as follows:

```
PARALLEL = 50;
```

where 50 is the number of random data sets that is drawn.

REPSE

The REPSE option is used to specify the resampling method that was used to create existing replicate weights or will be used to generate replicate weights (Fay, 1989; Korn & Graubard, 1999; Lohr, 1999; Asparouhov, 2009). Replicate weights are used in the estimation of standard errors of parameter estimates. The REPSE option has six settings: BOOTSTRAP, JACKKNIFE, JACKKNIFE1, JACKKNIFE2, BRR and FAY. There is no default. The REPSE option must be specified when replicate weights are used or generated.

With the BOOTSTRAP setting, the BOOTSTRAP option of the ANALYSIS command is used to specify the number of bootstrap draws used in the generation of the replicate weights. With the JACKKNIFE setting, the number of Jackknife draws is equal to the number of PSU's in the sample. A multiplier file is required for JACKKNIFE when replicate weights are used. The size of this file is one column with rows equal to the number of PSU's. For each PSU in a stratum, the value in the file is equal to the number of PSU's in the stratum minus one divided by the number of PSU's in the stratum. All PSU's in a stratum have the same value. If replicate weights are generated using JACKKNIFE, a multiplier file can be saved. JACKKNIFE1 cannot be used when data are stratified. JACKKNIFE2, balanced repeated replication (BRR), and FAY are available only when there are two PSU's in each stratum. The BRR and FAY resampling methods use Hadamard matrices. With BRR and FAY, the number of replicate weights is equal to the size of the Hadamard matrix. The REPSE option is specified as follows:

```
REPSE = BRR;
```

where BRR specifies that the balanced repeated replication resampling method is used to generate replicate weights.

For the FAY resampling method, a constant can be given that is used to modify the sample weights. The constant must range between zero and one. The default is .3. The REPSE option for the FAY setting is specified as follows:

```
REPSE = FAY (.5);
```

where .5 is the constant used to modify the sample weights.

BASEHAZARD

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

```
BASEHAZARD = ON;
```

With TYPE=MIXTURE, the ON and OFF settings have two alternatives, EQUAL and UNEQUAL. EQUAL is the default. With EQUAL, the baseline hazard parameters are held equal across classes. With BASEHAZARD=OFF, the baseline hazard parameters are held equal across classes as the default in line with Larsen (2004). To relax this equality, specify:

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

or

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

In continuous-time survival modeling, there are as many baseline hazard parameters as there are time intervals plus one. When the BASEHAZARD option of the ANALYSIS command is ON, these

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

[t];

where t is the time-to-event variable.

CHOLESKY

The CHOLESKY option is used in conjunction with ALGORITHM=INTEGRATION to decompose the continuous latent variable covariance matrix and the observed variable residual covariance matrix into orthogonal components in order to improve the optimization. The optimization algorithm starts out with Fisher Scoring used in combination with EM. The CHOLESKY option has two settings: ON and OFF. The default when all dependent variables are censored, categorical, and counts is ON except for categorical dependent variables when LINK=PROBIT. Then and in all other cases, it is OFF. To turn the CHOLESKY option ON, specify:

CHOLESKY = ON;

ALGORITHM

The ALGORITHM option is used in conjunction with TYPE=MIXTURE, TYPE=RANDOM, and TYPE=TWOLEVEL with maximum likelihood estimation to indicate the optimization method to use to obtain maximum likelihood estimates and to specify whether the computations require numerical integration. The ALGORITHM option is used with TYPE=TWOLEVEL and weighted least squares estimation to indicate the optimization method to use to obtain sample statistics for model estimation. There are four settings related to the optimization method: EM, EMA, FS, and ODLL. The default depends on the analysis type.

EM optimizes the complete-data loglikelihood using the expectation maximization (EM) algorithm (Dempster et al., 1977). EMA is an accelerated EM procedure that uses Quasi-Newton and Fisher Scoring optimization steps when needed. FS is Fisher Scoring. ODLL optimizes the observed-data loglikelihood directly.

To select the EM algorithm, specify the following:

```
ALGORITHM = EM;
```

The INTEGRATION setting of the ALGORITHM option is used in conjunction with numerical integration and the INTEGRATION option of the ANALYSIS command.

To select INTEGRATION, specify the following:

```
ALGORITHM = INTEGRATION;
```

The ALGORITHM option can specify an optimization setting in addition to the INTEGRATION setting, for example,

```
ALGORITHM = INTEGRATION EM;
```

OPTIONS RELATED TO NUMERICAL INTEGRATION

INTEGRATION

The INTEGRATION option is used to specify the type of numerical integration and the number of integration points to be used in the computation when ALGORITHM=INTEGRATION is used. The INTEGRATION option has three settings: STANDARD, GAUSSHERMITE, and MONTECARLO. The default is STANDARD. STANDARD uses rectangular (trapezoid) numerical integration. The default for TYPE=EFA and TYPE=TWOLEVEL with weighted least squares estimation is 7 integration points per dimension. For all other analyses, the default is 15 integration points per dimension. GAUSSHERMITE uses Gauss-Hermite integration with a default of 15 integration point per dimension. MONTECARLO uses randomly

generated integration points. The default number of integration points varies depending on the analysis type. In most cases, it is 500.

Following is an example of how the INTEGRATION option is used to change the number of integration points for the default setting of STANDARD.

```
INTEGRATION = 10;
```

where 10 is the number of integration points per dimension to be used in the computation. An alternative specification is:

```
INTEGRATION = STANDARD (10);
```

To select the MONTECARLO setting, specify:

```
INTEGRATION = MONTECARLO;
```

The default number of integration points varies depending on the analysis type. In most cases, 5000 integration points are used. Following is an example of how to specify a specific number of Monte Carlo integration points:

```
INTEGRATION = MONTECARLO (1000);
```

MCSEED

The MCSEED option is used to specify a random seed when the MONTECARLO setting of the INTEGRATION option is used. It is specified as follows:

```
MCSEED = 23456;
```

ADAPTIVE

The ADAPTIVE option is used to customize the numerical integration points for each observation during the computation. The ADAPTIVE option is available for each of the three settings of the INTEGRATION option. The ADAPTIVE option has two settings: ON and OFF. The default is ON. To turn the ADAPTIVE option off, specify:

ADAPTIVE = OFF;

INFORMATION

The **INFORMATION** option is used to select the estimator of the information matrix to be used in computing standard errors when the ML or MLR estimators are used for analysis. The **INFORMATION** option has three settings: **OBSERVED**, **EXPECTED**, and **COMBINATION**. **OBSERVED** estimates the information matrix using observed second-order derivatives; **EXPECTED** estimates the information matrix using expected second-order derivatives; and **COMBINATION** estimates the information matrix using a combination of observed and expected second-order derivatives. For MLR, **OBSERVED**, **EXPECTED**, and **COMBINATION** refer to the outside matrices of the sandwich estimator used to compute standard errors. The **INFORMATION** option is specified as follows:

INFORMATION = COMBINATION;

The default is to estimate models under missing data theory using all available data. In this case, the observed information matrix is used (Kenward & Molenberghs, 1998). For models with all continuous outcomes that are estimated without numerical integration, the expected information matrix is also available. For other outcome types and models that are estimated with numerical integration, the combination information matrix is also available.

BOOTSTRAP

The **BOOTSTRAP** option is used to request bootstrapping and to specify the type of bootstrapping and the number of bootstrap draws to be used in the computation. Two types of bootstrapping are available, standard non-parametric and residual parametric (Bollen & Stine, 1992; Efron & Tibshirani, 1993; Enders, 2002). Residual parametric bootstrap is the Bollen-Stine bootstrap. The **BOOTSTRAP** option requires individual data.

Standard non-parametric bootstrapping is available for the ML, WLS, WLSM, WLSMV, ULS, and GLS estimators. The reason that it is not available for MLR, MLF, MLM, and MLMV is that parameter estimates for these estimators do not differ from those of ML. Standard non-

parametric bootstrapping is not available for TYPE=EFA, COMPLEX, TWOLEVEL, THREELEVEL, CROSSCLASSIFIED, and RANDOM without ALGORITHM=INTEGRATION.

Residual parametric bootstrapping is available for only continuous outcomes using maximum likelihood estimation. In addition to the restrictions for standard non-parametric bootstrapping listed above, residual parametric bootstrapping is not available for TYPE=MIXTURE.

When the BOOTSTRAP option is used alone, bootstrap standard errors of the model parameter estimates are obtained for standard bootstrapping and bootstrap standard errors of the model parameter estimates and the chi-square p-value are obtained for residual bootstrapping. When the BOOTSTRAP option is used in conjunction with the CINTERVAL option of the OUTPUT command, bootstrap standard errors of the model parameter estimates and either symmetric, bootstrap, or bias-corrected bootstrap confidence intervals for the model parameter estimates can be obtained. The BOOTSTRAP option can be used in conjunction with the MODEL INDIRECT command to obtain bootstrap standard errors for indirect effects. When both MODEL INDIRECT and CINTERVAL are used, bootstrap standard errors and either symmetric, bootstrap, or bias-corrected bootstrap confidence intervals are obtained for the indirect effects.

The BOOTSTRAP option for standard bootstrapping is specified as follows:

```
BOOTSTRAP = 500;
```

where 500 is the number of bootstrap draws to be used in the computation. An alternative specification is:

```
BOOTSTRAP = 500 (STANDARD);
```

The BOOTSTRAP option for residual bootstrapping is specified as follows:

```
BOOTSTRAP = 500 (RESIDUAL);
```

where 500 is the number of bootstrap draws to be used in the computation.

LRTBOOTSTRAP

The LRTBOOTSTRAP option is used in conjunction with the TECH14 option of the OUTPUT command to specify the number of bootstrap draws to be used in estimating the p-value of the parametric bootstrapped likelihood ratio test (McLachlan & Peel, 2000). The default number of bootstrap draws is determined by the program using a sequential method in which the number of draws varies from 2 to 100. The LRTBOOTSTRAP option is used to override this default.

The LRTBOOTSTRAP option is specified as follows:

```
LRTBOOTSTRAP = 100;
```

where 100 is the number of bootstrap draws to be used in estimating the p-value of the parametric bootstrapped likelihood ratio test.

OPTIONS RELATED TO RANDOM STARTS

For TYPE=MIXTURE, TYPE=TWOLEVEL with categorical outcomes and weighted least squares estimation, and TYPE=EFA, random sets of starting values can be generated. Random starts can be turned off or done more thoroughly using the following set of options.

When TYPE=MIXTURE is used, random sets of starting values are generated as the default for all parameters in the model except variances and covariances. These random sets of starting values are random perturbations of either user-specified starting values or default starting values produced by the program. Maximum likelihood optimization is done in two stages. In the initial stage, random sets of starting values are generated. An optimization is carried out for ten iterations using each of the random sets of starting values. The ending values from the optimizations with the two highest loglikelihoods are used as the starting values in the final stage optimizations which are carried out using the default optimization settings for TYPE=MIXTURE.

When TYPE=TWOLEVEL with categorical outcomes and weighted least squares estimation or TYPE=EFA is used random sets of starting values are generated for the factor loading parameters in the model. For TYPE=TWOLEVEL with categorical outcomes and weighted least squares estimation, these random sets of starting values are random

perturbations of either user-specified starting values or default starting values produced by the program. For TYPE=EFA, these random sets of starting values are random perturbations of default starting values produced by the program.

STARTS

The STARTS option is used to specify the number of random sets of starting values to generate in the initial stage and the number of optimizations to use in the final stage. For TYPE=MIXTURE, the default is 20 random sets of starting values in the initial stage and 4 optimizations in the final stage. To turn off random starts, the STARTS option is specified as follows:

```
STARTS = 0;
```

Following is an example of how to use the STARTS option for TYPE=MIXTURE:

```
STARTS = 100 20;
```

specifies that 100 random sets of starting values are generated in the initial stage and 20 optimizations are carried out in the final stage using the default optimization settings for TYPE=MIXTURE.

Following are recommendations for a more thorough investigation of multiple solutions:

```
STARTS = 400 100;
```

or

```
STARTS = 1000 250;
```

For TYPE=EFA; TYPE=GENERAL; and TYPE=TWOLEVEL using the WLS, WLSM, WLSMV, and ULSVM estimators, the STARTS option is specified as follows:

```
STARTS = 10;
```

which specifies that 10 random sets of starting values are generated and ten optimizations are carried out.

STITERATIONS

The STITERATIONS option is used to specify the maximum number of iterations allowed in the initial stage. The default number of iterations is 10. For a more thorough investigation, 20 iterations can be requested as follows:

STITERATIONS = 20;

STCONVERGENCE

The STCONVERGENCE option is used to specify the value of the derivative convergence criterion to be used in the initial stage optimization. The default is one.

STSCALE

The STSCALE option is used to specify the scale of the random perturbation. The default is five which represents a medium level scale of perturbation.

STSEED

The STSEED option is used to specify the random seed for generating the random starts. The default value is zero.

OPTSEED

The OPTSEED option is used to specify the random seed that has been found to result in the highest loglikelihood in a previous analysis. The OPTSEED option results in no random starts being used.

K-1STARTS

The K-1STARTS option is used in conjunction with the TECH11 and TECH14 options of the OUTPUT command to specify the number of random sets of starting values to use in the initial stage and the number

of optimizations to use in the final stage for the k-1 class analysis model. When the OPTSEED option is used, the default is 20 random sets of starting values in the initial stage and 4 optimizations in the final stage. When the OPTSEED option is not used, the default is the same as what is used for the STARTS option. Following is an example of how to specify the K-1STARTS option:

```
K-1STARTS = 80 16;
```

which specifies that 80 random sets of starting values are generated in the initial stage and 16 optimizations are carried out in the final stage using the default optimization settings for TYPE=MIXTURE.

LRTSTARTS

The LRTSTARTS option is used in conjunction with the TECH14 option of the OUTPUT command to specify the number of starting values to use in the initial stage and the number of optimizations to use in the final stage for the k-1 and k class models when the data generated by bootstrap draws are analyzed. The default for the k-1 class model is 0 random sets of starting values in the initial stage and 0 optimizations in the final stage. One optimization is carried out for the unperturbed set of starting values. The default for the k class model is 40 random sets of starting values in the initial stage and 8 optimizations in the final stage.

Following is an example of how to use the LRTSTARTS option:

```
LRTSTARTS = 2 1 80 16;
```

which specifies that for the k-1 class model 2 random sets of starting values are used in the initial stage and 1 optimization is carried out in the final stage and for the k class model 80 random sets of starting values are used in the initial stage and 16 optimizations are carried out in the final stage.

RSTARTS

The RSTARTS option is used to specify the number of random sets of starting values to use for the GPA rotation algorithm and the number of rotated factor solutions with the best unique rotation function values to print for exploratory factor analysis. The default is 30 random sets of

starting values and printing of the best solution. Following is an example of how to use the RSTARTS option.

```
RSTARTS = 10 2;
```

which specifies that 10 random sets of starting values are used for the rotations and that the rotated factor solutions with the two best rotation function values will be printed.

ASTARTS

The ASTARTS option is used to specify the number of random sets of starting values to use for the alignment optimization. The default is 30.

H1STARTS

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

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

```
H1STARTS = 100 20;
```

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

DIFFTEST

The DIFFTEST option is used to obtain a correct chi-square difference test when the MLMV and the WLSMV estimators are used because the difference in chi-square values for two nested models using the MLMV or WLSMV chi-square values is not distributed as chi-square. The chi-square difference test compares the H0 analysis model to a less restrictive H1 alternative model in which the H0 model is nested. To obtain a correct chi-square difference test for MLMV or WLSMV, a

two-step procedure is needed. In the first step, the H1 model is estimated. In the H1 analysis, the DIFFTEST option of the SAVEDATA command is used to save the derivatives needed for the chi-square difference test. In the second step, the H0 model is estimated and the chi-square difference test is computed using the derivatives from the H0 and H1 analyses. The DIFFTEST option of the ANALYSIS command is used as follows to specify the name of the data set that contains the derivatives from the H1 analysis:

```
DIFFTEST = deriv.dat;
```

where deriv.dat is the name of the data set that contains the derivatives from the H1 analysis that were saved using the DIFFTEST option of the SAVEDATA command when the H1 model was estimated.

MULTIPLIER

The MULTIPLIER option is used with the JACKKNIFE setting of the RESPE option when replicate weights are used in the analysis to provide multiplier values needed for the computation of standard errors. The MULTIPLIER option is specified as follows:

```
MULTIPLIER = multiplier.dat;
```

where multiplier.dat is the name of the data set that contains the multiplier values needed for the computation of standard errors.

COVERAGE

The COVERAGE option is used with missing data to specify the minimum acceptable covariance coverage value for the unrestricted H1 model. The default value is .10 which means that if all variables and pairs of variables have data for at least ten percent of the sample, the model will be estimated. Following is an example of how to use the COVERAGE option:

```
COVERAGE = .05;
```

where .05 is the minimum acceptable covariance coverage value.

ADDFREQUENCY

The ADDFREQUENCY option is used to specify a value that is divided by the sample size and added to each cell with zero frequency in the two-way tables that are used in categorical data analysis. As the default, 0.5 divided by the sample size is added to each cell with zero frequency. The ADDFREQUENCY option is specified as follows:

ADDFREQUENCY = 0;

where the value 0 specifies that nothing is added to each cell with zero frequency. Any non-negative value can be used with this option.

OPTIONS RELATED TO ITERATIONS

ITERATIONS

The ITERATIONS option is used to specify the maximum number of iterations for the Quasi-Newton algorithm for continuous outcomes. The default number of iterations is 1,000.

SDITERATIONS

The SDITERATIONS option is used to specify the maximum number of steepest descent iterations for the Quasi-Newton algorithm for continuous outcomes. The default number of iterations is 20.

H1ITERATIONS

The H1ITERATIONS option is used to specify the maximum number of iterations for the EM algorithm for the estimation of the unrestricted H1 model. The default number of iterations is 2000.

MITERATIONS

The MITERATIONS option is used to specify the number of iterations allowed for the EM algorithm. The default number of iterations is 500.

MCITERATIONS

The MCITERATIONS option is used to specify the number of iterations for the M step of the EM algorithm for categorical latent variables. The default number of iterations is 1.

MUITERATIONS

The MUITERATIONS option is used to specify the number of iterations for the M step of the EM algorithm for censored, categorical, and count outcomes. The default number of iterations is 1.

RITERATIONS

The RITERATIONS option is used to specify the maximum number of iterations in the GPA rotation algorithm for exploratory factor analysis. The default number of iterations is 10000.

AITERATIONS

The AITERATIONS option is used to specify the maximum number of iterations in the alignment optimization. The default is 5000.

OPTIONS RELATED TO CONVERGENCE**CONVERGENCE**

The CONVERGENCE option is used to specify the value of the derivative convergence criterion to be used for the Quasi-Newton algorithm for continuous outcomes. The default convergence criterion for TYPE=TWOLEVEL, TYPE=MIXTURE, TYPE=RANDOM, and ALGORITHM=INTEGRATION is .000001. The default convergence criterion for all other models is .00005.

H1CONVERGENCE

The H1CONVERGENCE option is used to specify the value of the convergence criterion to be used for the EM algorithm for the estimation of the unrestricted H1 model. The default convergence criterion for

TYPE=THREELEVEL and TYPE=CROSSCLASSIFIED is .001. The default convergence criterion for all other models is .0001.

LOGCRITERION

The LOGCRITERION option is used to specify the absolute observed-data loglikelihood change convergence criterion for the EM algorithm. The default convergence criterion for TYPE=TWOLEVEL, TYPE=RANDOM, and ALGORITHM=INTEGRATION is .001. The default convergence criterion for TYPE=MIXTURE with PARAMETERIZATION=PROBABILITY is .0001. The default convergence criterion for all other models is .0000001.

RLOGCRITERION

The RLOGCRITERION option is used to specify the relative observed-data loglikelihood change convergence criterion for the EM algorithm. The default convergence criterion for TYPE=TWOLEVEL, TYPE=RANDOM, and ALGORITHM=INTEGRATION is .000001. The default convergence criterion for all other models is .0000001.

MCONVERGENCE

The MCONVERGENCE option is used to specify the observed-data log likelihood derivative convergence criterion for the EM algorithm. The default convergence criterion for TYPE=TWOLEVEL, TYPE=RANDOM, and ALGORITHM=INTEGRATION is .001. The default for TYPE=MIXTURE with PARAMETERIZATION=PROBABILITY is .0001. The default convergence criterion for all other models is .000001.

MCCONVERGENCE

The MCCONVERGENCE option is used to specify the complete-data log likelihood derivative convergence criterion for the M step of the EM algorithm for categorical latent variables. The default convergence criterion is .000001.

MUCONVERGENCE

The MUCONVERGENCE option is used to specify the complete-data log likelihood derivative convergence criterion for the M step of the EM algorithm for censored, categorical, and count outcomes. The default convergence criterion is .000001.

RCONVERGENCE

The RCONVERGENCE option is used to specify the convergence criterion for the GPA rotation algorithm for exploratory factor analysis. The default convergence criterion is .00001.

ACONVERGENCE

The ACONVERGENCE option is used to specify the convergence criterion for the derivatives of the alignment optimization. The default is 0.001.

MIXC

The MIXC option is used to specify whether to use the number of iterations or the convergence criterion to terminate the M step iterations of the EM algorithm for categorical latent variables. Following is an example of how to select the convergence criterion being fulfilled:

```
MIXC = CONVERGENCE;
```

MIXU

The MIXU option is used to specify whether to use the number of iterations or the convergence criterion to terminate the M step iterations of the EM algorithm for censored, categorical, and count outcomes. Following is an example of how to select the convergence criterion being fulfilled:

```
MIXU = CONVERGENCE;
```

LOGHIGH

The LOGHIGH option is used to specify the maximum value allowed for the logit thresholds of the latent class indicators. The default is +15.

LOGLOW

The LOGLOW option is used to specify the minimum value allowed for the logit thresholds of the latent class indicators. The default is -15.

UCELLSIZE

The UCELLSIZE option is used to specify the minimum expected cell size allowed for computing chi-square from the frequency table of the latent class indicators when the corresponding observed cell size is not zero. The default value is .01.

VARIANCE

The VARIANCE option is used in conjunction with TYPE=RANDOM and TYPE=TWOLEVEL when ESTIMATOR=ML, MLR, or MLF to specify the minimum value that is allowed in the estimation of the variance of the random effect variables and the variances of the between-level outcome variables. The default value is .0001.

SIMPLICITY

The SIMPLICITY option is used to select the simplicity criterion of the alignment optimization. The simplicity function is optimized at a solution with a few large non-invariant parameters and many invariant parameters rather than many medium-sized non-invariant parameters. The SIMPLICITY option has two settings: SQRT and FOURTHRT. SQRT is the default. The SQRT setting takes the square root of the weighted component loss function. The FOURTHRT setting takes the double square root of the weighted component loss function. It may in some cases further reduce small significant differences.

TOLERANCE

The TOLERANCE option is used to specify the simplicity tolerance value of the alignment optimization which must be positive. The default is 0.01.

METRIC

The METRIC option is used to specify the factor variance metric of the alignment optimization. The METRIC option has two settings: REFGROUP and PRODUCT. REFGROUP is the default where the factor variance is fixed at one in the reference group. The PRODUCT setting sets the product of the factor variances in all of the groups to one. The PRODUCT setting is not allowed with ALIGNMENT=FIXED.

MATRIX

The MATRIX option identifies the matrix to be analyzed. The default for continuous outcomes is to analyze the covariance matrix. The following statement requests that a correlation matrix be analyzed:

```
MATRIX = CORRELATION;
```

The analysis of the correlation matrix is allowed only when all dependent variables are continuous and there is a single group analysis with no mean structure. Only the WLS estimator is allowed for this type of analysis.

For models with all categorical dependent variables, the correlation matrix is always analyzed. For models with combinations of categorical and continuous dependent variables, the variances for the continuous dependent variables are always included.

OPTIONS RELATED TO BAYES ESTIMATION AND MULTIPLE IMPUTATION

POINT

The POINT option is used to specify the type of Bayes point estimate to compute. The POINT option has three settings: MEDIAN, MEAN, and

MODE. The default is **MEDIAN**. With the **MODE** setting, the mode reported refers to the multivariate mode of the posterior distribution. This mode is different from the univariate mode reported in the plot of the Bayesian posterior parameter distribution. To request that the mean be computed, specify:

```
POINT = MEAN;
```

CHAINS

The **CHAINS** option is used to specify how many independent Markov chain Monte Carlo (MCMC) chains to use. The default is two. To request that four chains be used, specify:

```
CHAINS = 4;
```

With multiple chains, parallel computing uses one chain per processor. To benefit from this speed advantage, it is important to specify the number of processors using the **PROCESSORS** option.

BSEED

The **BSEED** option is used to specify the seed to use for random number generation in the Markov chain Monte Carlo (MCMC) chains. The default is zero. If one chain is used, the seed is used for this chain. If more than one chain is used, the seed is used for the first chain and is the basis for generating seeds for the other chains. The randomly generated seeds for the other chains can be found in **TECH8**. If the same seed is used in a subsequent analysis, the other chains will have the same seeds as in the previous analysis. To request a seed other than zero be used, specify:

```
BSEED = 5437;
```

STVALUES

The **STVALUES** option is used to specify starting value information (Asparouhov & Muthén, 2010b). The **STVALUES** option has three settings: **UNPERTURBED**, **PERTURBED**, and **ML**. The default is **UNPERTURBED**. If the **UNPERTURBED** setting is specified, the default or user-specified starting values are used. If the **PERTURBED**

setting is used, a BSEED value must be specified. The default or user-specified starting values are randomly perturbed using the BSEED value. If the ML setting is used, the model is first estimated using maximum likelihood estimation and the maximum likelihood parameter estimates are used as starting values in the Bayesian analysis. To request that maximum likelihood parameters be used as starting values, specify:

```
STVALUES = ML;
```

PREDICTOR

The PREDICTOR option is used with ESTIMATOR=BAYES to specify how a categorical mediator variable is treated when it is an independent variable in a regression and how an observed exogenous binary variable is treated when it is brought into the model and put on the CATEGORICAL list. The PREDICTOR option has two settings: LATENT and OBSERVED. The default is LATENT where the predictor variable is treated as a continuous latent response variable underlying the categorical variable. When the OBSERVED setting is specified, the predictor variable is treated as a continuous observed variable. Muthén, Muthén, and Asparouhov (2016, section 9.8.2) recommend using the OBSERVED setting for an observed exogenous binary variable on the CATEGORICAL list. To request that the predictor variable be treated as a continuous observed variable, specify:

```
PREDICTOR = OBSERVED;
```

ALGORITHM

The ALGORITHM option is used to specify the Markov chain Monte Carlo (MCMC) algorithm to use for generating the posterior distribution of the parameters (Gelman et al., 2004). The ALGORITHM option has two settings: GIBBS and MH. The default is GIBBS. The GIBBS setting uses the Gibbs sampler algorithm which divides the parameters and the latent variables into groups that are conditionally and sequentially generated. The GIBBS setting has four choices: PX1, PX2, PX3, and RW. The default is PX1. PX1, PX2, and PX3 use parameter extension techniques to generate correlation and covariance matrices. PX1 is described in Asparouhov and Muthén (2010b). PX2 is described in Boscardin et al. (2008). PX3 is described in Liu and Daniels (2006). RW uses a random walk, Metropolis-Hastings algorithm to generate

correlation and covariance matrices (Chib & Greenberg, 1998). This algorithm can generate a covariance matrix with an arbitrary structure. Following is an example of how to select an alternative choice for the GIBBS setting:

```
ALGORITHM = GIBBS (PX3);
```

The MH setting uses the Metropolis-Hastings algorithm to generate all of the parameters simultaneously using the observed-data loglikelihood. The MH setting uses maximum likelihood starting values. The MH proposal distribution uses the estimated covariance matrix of the maximum likelihood parameter estimates. The MH algorithm is not available for TYPE=MIXTURE or TYPE=TWOLEVEL. To request that the Metropolis-Hastings algorithm be used, specify:

```
ALGORITHM = MH;
```

BCONVERGENCE

The BCONVERGENCE option is used to specify the value of the convergence criterion to use for determining convergence of the Bayesian estimation using the Gelman-Rubin convergence criterion (Gelman & Rubin, 1992). The Gelman-Rubin convergence criterion determines convergence by considering within and between chain variability of the parameter estimates in terms of the potential scale reduction (PSR) to determine convergence (Gelman et al., 2004, pp. 296-298). The default is 0.05. The BCONVERGENCE value is used in the following formula (Asparouhov & Muthén, 2010b):

$$a = 1 + \text{BCONVERGENCE} * \text{factor},$$

such that convergence is obtained when $\text{PSR} < a$ for each parameter. The factor value ranges between one and two depending on the number of parameters. With one parameter, the value of factor is one and the value of a is 1.05 using the default value of BCONVERGENCE. With a large number of parameters, the value of factor is 2 and the value of a is 1.1 using the default value of BCONVERGENCE.

With a single chain, PSR is defined using the third and the fourth quarters of the chain. The first half of the chain is discarded as a burnin phase. To request a stricter convergence criterion, specify:

BCONVERGENCE = .01;

BITERATIONS

The BITERATIONS option is used to specify the maximum and minimum numbers of iterations for each Markov chain Monte Carlo (MCMC) chain when the potential scale reduction (PSR) convergence criterion (Gelman & Rubin, 1992) is used. The default for the maximum number of iterations is 50,000. The default for the minimum number of iterations is zero. To request more Bayes iterations for each chain, specify:

BITERATIONS = 60000 (2000);

where 60,000 is the maximum number of iterations and 2,000 is the minimum number of iterations when the PSR convergence criterion (Gelman & Rubin, 1992) is used.

Another specification is:

BITERATIONS = (2000);

where the default of 50,000 is the maximum number of iterations and 2,000 is the minimum number of iterations when the PSR convergence criterion is used.

FBITERATIONS

The FBITERATIONS option is used to specify a fixed number of iterations for each Markov Chain Monte Carlo (MCMC) chain when the potential scale reduction (PSR) convergence criterion (Gelman & Rubin, 1992) is not used. There is no default. When using this option, it is important to use other means to determine convergence such as posterior parameter trace plots. To request a fixed number of iterations for each Markov Chain Monte Carlo (MCMC) chain, specify:

FBITERATIONS = 30000;

THIN

The THIN option is used to specify which iterations from the posterior distribution to use in the parameter estimation. When a chain is mixing poorly with high auto-correlations, the estimation can be based on every k-th iteration rather than every iteration. This is referred to as thinning. The default is 1 in which case every iteration is used. To request that every 20th iteration be used, specify:

```
THIN = 20;
```

which means that the first iteration used is 20, the second is 40, the third is 60 etc.

MDITERATIONS

The MDITERATIONS option is used with the MODE setting of the POINT option to specify the maximum number of iterations to use to compute the multivariate mode in Bayes estimation. The default is 10,000. If the number of iterations used in the estimation exceeds the number of iterations specified using the MDITERATIONS option, the number of iterations specified using the MDITERATIONS option is used. This number of iterations is selected from the total iterations using equally-spaced intervals. To request that more iterations be used to compute the multivariate mode, specify:

```
MDITERATIONS = 15000;
```

KOLMOGOROV

The KOLMOGOROV option is used to request a Kolmogorov-Smirnov test of equality of the posterior parameter distributions across the different chains using draws from the chains. The default is 100. To request more draws, specify:

```
KOLMOGOROV = 1000;
```

PRIOR

The PRIOR option is used to request a plot of the prior distribution for each parameter that has a proper prior. The plot of the prior

distributions can be viewed by choosing Bayesian prior distributions from the Plot menu of the Mplus Editor. The default is 1,000 draws from the prior distribution. To request more draws, specify:

```
PRIOR = 5000;
```

INTERACTIVE

The INTERACTIVE option is used to allow changes in technical specifications during the iterations of an analysis when TECH8 is used. This is useful in analyses that are computationally demanding. If a starting value set has computational difficulties, it can be skipped. If too many random starts have been chosen, the STARTS option can be changed. If a too strict convergence criterion has been chosen, the MCONVERGENCE option can be changed. Following is an example of how to use the INTERACTIVE option:

```
INTERACTIVE = control.dat;
```

where control.dat is the name of the file that contains the technical specifications that can be changed during an analysis. This file is created automatically and resides in the same directory as the input file. The following options of the ANALYSIS command are contained in this file: STARTS, MITERATIONS, MCONVERGENCE, LOGCRITERION, and RLOGCRITERION. No other options can be used in this file except the INTERRUPT statement which is used to skip the current starting value set and go to the next starting value set. It has settings of 0 and 1. A setting of 0 specifies that a starting value set is not skipped. A setting of 1 specifies that the starting value set is skipped. As the default, the INTERRUPT statement is set to 0 and the other options are set to either the program default values or the values specified in the input file.

The following file is automatically created and given the name specified using the INTERACTIVE option.

```
INTERRUPT = 0  
STARTS = 200 50  
MITERATIONS = 500  
MCONVERGENCE = 1.0E-06  
LOGCRITERION = 1.0E-003
```

```
RLOGCRITERION = 1.0E-006
```

When the file is modified and saved, the new settings go into effect immediately and are applied at each iteration. Following is an example of a modified control.dat file where INTERRUPT and STARTS are changed:

```
INTERRUPT = 1  
STARTS = 150 50  
MITERATIONS = 500  
MCONVERGENCE = 1.0E-06  
LOGCRITERION = 1.0E-003  
RLOGCRITERION = 1.0E-006
```

PROCESSORS

The PROCESSORS option is used to specify the number of processors to be used for parallel computing to increase computational speed. When random starts are used, the PROCESSORS option is used in conjunction with the STARTS option to determine the number of threads to be used for parallel computing. The default is one processor and one thread. Parallel computing is not available for all analyses. For some analyses, multiple processors are used alone. In other analyses, multiple processors are used together with threads.

MULTIPLE PROCESSORS

The use of multiple processors without threads is available for TYPE=MIXTURE; Bayesian analysis with more than one chain unless STVALUES=ML; models that require numerical integration; models with all continuous variables, missing data, and maximum likelihood estimation; and TYPE=TWOLEVEL with categorical outcomes and ESTIMATOR= WLSMV. In these cases, the PROCESSORS option is specified using one number as shown below:

```
PROCESSORS = 8;
```

where 8 is the number of processors to be used for parallel computing.

For Bayesian analysis, the PROCESSORS option is specified using one number as shown below:

```
PROCESSORS = 2;
```

where 2 is the number of processors used for parallel computing. The number of processors used cannot exceed the number of chains. If there are more processors than chains, only the number of processors equal to the number of chains is used. If there are more chains than processors, each processor carries out one chain until it is completed and then the remaining chains are carried out.

PROCESSORS AND THREADS

When processor and threads are used together, the threads are distributed across the processors and the memory used is a multiple of the number of threads. For large models that require a lot of memory, it is important to have fewer threads than processors because computations are slower or impossible when the memory used by all processors exceeds the memory limit.

The use of multiple processors and multiple threads with random starts as the default is available for `TYPE=MIXTURE`; Bayesian analysis with more than one chain if `STVALUES=ML`; and models that require numerical integration. They are also available for `TYPE=RANDOM` and `TYPE=TWOLEVEL` and `THREELEVEL` with continuous outcomes using `ESTIMATOR=ML`, `MLR`, and `MLF` without numerical integration if the `STARTS` option is used. Without random starts only one processor is used in these cases.

When the `PROCESSORS` option is used with random starts, it is used with two numbers, the number of processors and the number of threads. The number of threads is the smaller of the number of threads specified using the `PROCESSORS` option or the number of final perturbations specified using the `STARTS` option. Following is an example:

```
PROCESSORS = 8 4;  
STARTS = 400 40;
```

where 4 is the number of threads and 40 is the number of final stage optimizations. Because four is smaller than 40, the number of threads for this example is four. Two processors are distributed across each of the four threads. Each of the four threads carries out 100 initial stage

CHAPTER 16

and ten final stage optimizations. When a thread completes, its processors are distributed across the remaining threads.

If the number of threads is not specified, it is the same as the number of processors. In this case, the `PROCESSORS` option is specified as follows:

```
PROCESSORS = 4;
```

where 4 is the number of processors to be used in the analysis for parallel computing. The number of threads is also 4.