VERSION 9 MPLUS LANGUAGE ADDENDUM

In this addendum, changes introduced in Version 9 are described. They include new features and corrections to minor problems that have been found since the release of Version 8.11 in May 2024. Three new web talks are available. Web Talk 8 describes how to carry out multistep mixture modeling for models with one categorical latent variable like LCA. Web Talk 9 describes a new method to analyze LCA with direct effects using PSEM. Web Talk 10 describes how to carry out multistep mixture modeling for models with more than one categorical latent variable like LTA.

Following are the new features in Version 9:

- Penalized Structural Equation Modeling (PSEM) has been extended (Asparouhov & Muthen, 2025a). For continuous variables, PSEM is now available for TYPE=TWOLEVEL and TYPE=THREELEVEL. For categorical, censored, count, nominal, and continuous-time survival variables, PSEM is available for single-level and TYPE=TWOLEVEL. For categorical variables using WLSMV, PSEM is available for single-level and TYPE=TWOLEVEL. PSEM is available for TYPE=MIXTURE and TYPE=TWOLEVEL MIXTURE. Web Talk 9 discusses applications of PSEM to LCA with direct effects.
- Latent variable decomposition (latent variable centering) of observed exogenous independent variables with random slopes for TYPE=THREELEVEL with ESTIMATOR=BAYES (Asparouhov & Muthen, 2025b)
- DSEM for TYPE=THREELEVEL with continuous variables (Asparouhov & Muthen, 2025b). This is a first, experimental version of DSEM3.
- Settings for the AUXILIARY option and multistep mixture modeling extended. For applications, see Web Talk 8.
- TYPE=IMPUTATION added for multistep mixture modeling. See Web Talk 8.
- Output added for multistep mixture modeling. For examples, see Web Talks 8, 9, and 10.
- NAMES option added to DATA IMPUTATION. For applications, see Web Talk 8.
- BOOTSTRAP option added for TYPE=TWOLEVEL (Asparouhov & Muthen, 2025c)
- ESTIMATOR=BAYES for censored variables for single-level and TYPE=TWOLEVEL including DSEM
- Mplus Editors's CALCULATOR available for models with one latent class variable such as LCA including the computation of probabilities for the latent class indicators as a function of covariates. For applications see Web Talk 9.
- Variance Inflation Factor (VIF) for information about collinearity among covariates
- Effective Sample Size (ESS) for Bayes. See the FAQ ESS – Effective sample size for Bayes.
- New choices for ALGORITHM=MH for CT-RDSEM (Asparouhov & Muthen, 2024)
- Partial correlation output (Asparouhov & Muthen, 2025d)
- MODEL TEST for TYPE=CROSSCLASSIFIED
- SAVE=CPROBABILITIES for DATA IMPUTATION and TYPE=IMPUTATION
- CONVERGENCE option added to the MONTECARLO command
- MODEL command extensions of the dot language for LTA. For applications, see Web Talk 10.

MULTISTEP APPROACHES WITH MIXTURE MODELING

The AUXILIARY option is used in conjunction with TYPE=MIXTURE to provide automatic estimation of the 3-step and BCH approaches to multistep mixture modeling. The automatic approach can be used in conjunction with TYPE=IMPUTATION. It is not allowed for models with more than one categorical latent variable or for models for which numerical integration is required. These models must use the manual approach (Asparouhov & Muthén, 2014, 2021, Web Talk 8).

The AUXILIARY option has six settings. One is for covariates used in the multinomial logistic regression for the categorical latent variable. The other five are for testing the differences between the means or probabilities of distal outcomes across classes.

The setting for covariates is R3STEP (Vermunt, 2010; Asparouhov & Muthén, 2014). It is used to identify a set of variables not used in the first step of the analysis that is used in the last step as covariates in a multinomial logistic regression for a categorical latent variable. The multinomial logistic regression uses all covariates jointly. Confidence intervals are given for the odds ratio of each covariate.

Of the five settings for distal outcomes, three use the 3-step method and two use the BCH method. A set of variables, often distal outcomes, not used in the first step of the analysis are identified. For continuous variables, the equality of means across classes is tested. For categorical variables, the equality of probabilities across classes is tested. These tests are done one variable at a time. Odds ratios are also provided for categorical variables. More than one of these five settings can be used in a single analysis.

The 3-step settings are D3STEP, DU3STEP, and D3STEPC. D3STEP (Asparouhov & Muthén, 2014) tests the equality of means for continuous variables across classes with the variances equal across classes. DU3STEP (Asparouhov & Muthén, 2014) tests the equality of means for continuous variables across classes with the variances unequal across classes. D3STEPC tests the equality of probabilities for categorical variables across classes. Odds ratios are also provided.

The BCH settings are BCH and BCHC. BCH (Vermunt, 2010; Bakk & Vermunt, 2016) tests the equality of means for continuous variables across classes with the variances equal across classes. BCHC tests the equality of probabilities for categorical variables across classes. Odds ratios are also provided.

All of the settings are specified in the same way. The setting in parentheses is placed behind each variable or list of variables on the AUXILIARY statement. Alternatively, the setting in parentheses can come first followed by variables and lists of variables. All of the variables on the AUXILIARY statement will be saved if the SAVEDATA command is used and will be available for plots if the PLOT command is used.

Following is an example of how to specify the R3STEP setting:

AUXILIARY = race (R3STEP) ses (R3STEP) x1-x5 (R3STEP);

where race, ses, x1, x2, x3, x4, and x5 will be used as covariates in a multinomial logistic regression.

An alternative specification is:

AUXILIARY = (R3STEP) race ses x1-x5;

where all variables and lists of variables after (R3STEP) will be used as covariates in a multinomial logistic regression. This is convenient when there are several variables that cannot be specified using the list function.

Following is an example of how to specify the settings for testing the equality of means and probabilities across classes. It is possible to use more than one of these settings in the same AUXILIARY statement:

AUXILIARY = abuse (BCH) dropout (BCHC) y1-y5 (D3STEP);

where the equality of the means of abuse, the probabilities of dropout, and the means of y1, y2, y3, y4, and y5 will be tested across classes.

NAMES

The NAMES option of the DATA IMPUTATION command is used to specify the names of the variables to use to impute missing values when variables beyond those listed in the USEVARIABLES list of the VARIABLE command are used.

ALGORITHM FOR BAYES

The ALGORITHM option for Bayes estimation 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 (2023). 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. The MH setting has six choices: MH1, UNIMH, MH2, MH3, MH4, and MH5. The default is MH1. MH1 uses multivariate updating within each block of parameters based on a random walk algorithm. UNIMH, MH2, MH3, MH4, and MH5 are for Continuous-time Dynamic Structural Equation Models (CT-DSEM, Asparouhov & Muthén, 2024). UNIMH uses univariate updating within each block of parameters based on a random walk algorithm. MH2 is described in Carlin and Louis (2009, p. 133). It uses multivariate updating within each block of parameters based on a proposal distribution for the variance block of parameters and a correlated random walk for the drift block of parameters. MH3 uses multivariate updating for the variance block of parameters based on a random walk and for the drift block of parameters using a correlated random walk. MH4 is described in Carlin and Louis (2009, p. 137). It uses multivariate updating within each block of parameters based on a Hamiltonian Montecarlo algorithm. MH5 uses a multivariate updating jointly for both the variance and drift blocks of parameters using a Hamiltonian Montecarlo algorithm.

To request that the Metropolis-Hastings algorithm be used, specify:

ALGORITHM = MH (MH1);

or

ALGORITHM = MH;

CONVERGENCE FOR MONTECARLO

The CONVERGENCE option of the MONTECARLO command is used to determine whether a replication is considered completed or not. The CONVERGENCE option has two settings: BASIC and STRICT. BASIC is the default. BASIC considers a converged replication completed even if it gets a non-positive definite warning. STRICT does not consider a converged replication completed if it gets a non-positive definite warning.

REFERENCES

Asparouhov, T. & Muthén, B. (2014). Auxiliary variables in mixture modeling: A 3-step approach using Mplus. Mplus Web Notes: No. 15. www.statmodel.com.

Asparouhov, T. & Muthén, B. (2021). Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model. Mplus Web Notes: No. 21. www.statmodel.com.

Asparouhov, T. & Muthén, B. (2023). Bayesian analysis using Mplus: Technical implementation. Technical Report. Los Angeles: Muthén & Muthén.

Asparouhov, T. & Muthén, B. (2024). Continuous time dynamic structural equation models. Technical Report. Los Angeles: Muthén & Muthén.

Asparouhov, T. & Muthén, B. (2025a). Methodological advances with penalized structural equation models. Structural Equation Modeling, 32 (4), 688-716.

Asparouhov, T. & Muthén, B. (2025b). Three-level dynamic modeling. Technical Report. Los Angeles: Muthén & Muthén.

Asparouhov, T. & Muthén, B. (2025c). Bootstrap in two-level models. Technical Report. Los Angeles: Muthén & Muthén.

Asparouhov, T. & Muthén, B. (2025d). Partial correlations. Mplus Web Notes: No. 26. www.statmodel.com.

Bakk, Z, & Vermunt, J.K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. Structural Equation Modeling, 23, 20-31.

Boscardin, J., Zhang, X., & Belin, T. (2008). Modeling a mixture of ordinal and continuous repeated measures. Journal of Statistical Computation and Simulation, 78, 873-886.

Carlin, B.P. & Louis, T.A. (2009). Bayesian methods for data analysis. Third edition. Boca Raton, FL: Chapman and Hall/CRC Press.

Chib, S. & Greenberg, E. (1998). Bayesian analysis of multivariate probit models. Biometrika, 85, 347-361.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). Bayesian data analysis. Second edition. New York: Chapman & Hall.

Liu, X. & Daniels, M.J. (2006). A new algorithm for simulating a correlation matrix based on parameter expansion and re-parameterization. Journal of Computational and Graphical Statistics, 15, 897-914.

Vermunt, J.K. (2010). Latent class modeling with covariates: Two improved three-step approaches. Political Analysis, 18, 450-469.