

ONLINE SUPPLEMENTAL MATERIAL

MIXTURE MODELING FOR ORGANIZATIONAL BEHAVIOR RESEARCH

ALEXANDRE J. S. MORIN, MATTHEW J. W. MCLARNON, AND DAVID LITALIEN

Appendix A. Technical Considerations in the Estimation of Mixture Models	S6
<i>Mixture Indicators</i>	S6
<i>Random Starts</i>	S8
<i>Class Enumeration</i>	S9
<i>References</i>	S11
Appendix B. Mplus Input Code for the Estimation of Mixture Models	S19
<i>Data</i>	S19
<i>Variable</i>	S19
<i>Analysis</i>	S21
<i>Model</i>	S24
<i>Output</i>	S24
<i>Latent Profile Analysis</i>	S25
<i>Latent Profile Analysis with Covariates</i>	S30
Predictors	S30
Outcomes	S34
Correlates	S38
<i>Procedure for the Manual Implementation of Auxiliary Approaches</i>	S38
BCH Approach	S39

Three-Step Approach	S42
<i>Factor Mixture Analysis</i>	S45
<i>Multiple-Group Latent Profile Analyses</i>	S46
Configural Similarity	S46
Structural Similarity	S49
Dispersion Similarity	S51
Distributional Similarity	S52
<i>Multiple-Group Latent Profile Analyses with Predictors</i>	S54
Relations Freely Estimated Across Subgroups	S54
Predictive Similarity	S56
<i>Multiple-Group Latent Profile Analyses with Outcomes</i>	S59
Relations Freely Estimated Across Subgroups	S59
Explanatory Similarity	S62
<i>Mixture Regression Analysis</i>	S65
<i>Multiple Group Mixture Regression Analysis</i>	S68
Configural Similarity	S68
Regression Similarity	S70
Structural Similarity	S71
Dispersion Similarity	S73
Distributional Similarity	S74
<i>Multiple Group Mixture Regression Analysis with Predictors</i>	S76
Relations Freely Estimated Across Subgroups	S76
Predictive Similarity	S78
<i>Multiple Group Mixture Regression Analysis with Outcomes</i>	S79
Relations Freely Estimated Across Subgroups	S79

Explanatory Similarity	S81
<i>Longitudinal Latent Profile Analysis</i>	S83
Configural Similarity	S83
Structural Similarity	S85
Dispersion Similarity	S86
Distributional Similarity	S86
<i>Latent Transition Analysis (from an LPA model of Dispersion Similarity)</i>	S88
<i>Latent Transition Analysis with Predictors</i>	S89
Relations Freely Estimated Across Profiles and Time Points	S89
Relations Freely Estimated Across Time Points	S92
Predictive Similarity	S93
<i>Latent Transition Analysis with Outcomes</i>	S95
Relations Freely Estimated Across Time Points	S95
Explanatory Similarity	S97
<i>Linear Growth Mixture Analyses</i>	S99
Alternative A – Homoscedastic Residuals (equal across time points)	S102
Alternative B – Class-Invariant Residuals	S103
Alternative C – Class-Invariant Latent Variance–Covariance Matrix	S103
Alternative D – Class-Invariant Homoscedastic Residuals (equal across time points)	S104
Alternative E – Class-Invariant Latent Variance–Covariance Matrix and Homoscedastic Residuals	S105
Alternative F – Class-Invariant Latent Variance–Covariance Matrix and Class-Invariant Residuals	S105
Alternative G – Class-Invariant Latent Variance–Covariance Matrix	

and Class-Invariant Homoscedastic Residuals	S106
<i>Quadratic Growth Mixture Analysis</i>	S107
<i>Piecewise Growth Mixture Analysis</i>	S108
<i>Latent Basis Growth Mixture Analysis</i>	S110
<i>Mehta and West (2000) Test</i>	S111
Initial Model	S112
Constrained Model	S112
<i>Age-Defined Linear Growth Mixture Analysis</i>	S112
<i>Age-Defined Quadratic Growth Mixture Analysis</i>	S115
<i>Growth Mixture Analysis: Time-Invariant Predictors</i>	S118
Null-Effects Model	S118
Alternative A – Freely Estimated Effects on Class Membership Only	S119
Alternative B – Effects on Class Membership and Class-Invariant Effects on the Growth Factors	S120
Alternative C – Effects on Class Membership and Class-Varying Effects on the Growth Factors	S121
<i>Growth Mixture Analysis: Time-Invariant Outcomes</i>	S122
<i>Growth Mixture Analysis: Time-Varying Predictors</i>	S124
Null-Effects Model	S124
Alternative A – Effects Free to Vary Across Time Points but Equal Across Profiles	S125
Alternative B – Effects Free to Vary Across Time Points and Profiles	S126
Alternative C – Effects Free to Vary Across Profiles but Not Time Points	S128
Alternative D – Effects Constrained to Equality Across Time Points and Profiles	S130

<i>Growth Mixture Analysis: Time-Varying Outcomes</i>	S131
Null-Effects Model	S132
Alternative A – Effects Free to Vary Across Time Points but Equal Across Profiles	S133
Alternative B – Effects Free to Vary Across Time Points and Profiles	S134
Alternative C – Effects Free to Vary Across Profiles but Not Time Points	S136
Alternative D – Effects Constrained to Equality Across Time Points and Profiles	S138

Appendix A

Technical Considerations in the Estimation of Mixture Models

MIXTURE INDICATORS

Most commonly mixture models are estimated on the basis of either continuous or categorical observed data, though other types of data such as count or nominal variables can also be used. When categorical data (i.e., binary or polytomous items) is used, the term *latent class analysis* (LCA) is often used to describe the mixture model. In LCA, the distinct subpopulations recovered differ on the basis of item response probabilities (e.g., McLarnon, DeLongchamp, & Schneider, 2019; Wang & Wang, 2012). Latent profile analysis (LPAs), on the other hand, are most often estimated using scale scores on the various indicators of interest (i.e., taking the average of the items used to assess a focal construct and using those scores as input data; see McLarnon, Carswell, & Schneider, 2015; O’Neill, McLarnon, Hoffart, Woodley, & Allen, 2018; O’Neill, McLarnon, Xiu, & Law, 2016; or by using factor scores exported from preliminary measurement models, see below). Throughout, we retain use of LPA given that the LCA–LPA distinction is somewhat subtle, and that categorical, ordinal, and continuous indicator variables can be included in the same mixture model (i.e., as a form of hybrid model; see Berlin, Williams, & Parra, 2014; Muthén & Muthén, 2017).

This supplemental material provides an overview of a number of nuanced issues associated with thoroughly conducting LPA, and is structured in the following way: first, we highlight the use of factor scores as the input variables for an LPA. Second, we discuss the technical issues of random starts in LPA. Third, we present a number of potential guidelines that researchers may be able to use to make decisions on model fit (i.e., profile or class enumeration). Finally, we provide a number of detailed *Mplus* syntax templates that researchers may be able to leverage to conduct LPAs and its variants.

Although using observed scale scores as input to an LPA is common, there is an

important limitation in doing so: scale scores generally have less-than-perfect reliability, and may therefore be impacted by measurement error. In contrast, a latent variable (stemming from a factor analytic model) naturally controls for measurement error (e.g., Bollen, 1989).

Unfortunately, applications of fully latent mixture models in which the means of latent factors (specified within a confirmatory factor model) vary across profiles are few (see, however, Morin, Scalas, & Marsh, 2015). This can, in part, be explained by the complexity of mixture models, which may make fully latent approaches computationally intractable with common computer hardware due to nonconvergence or convergence upon improper solutions.

A strategy that reaches a compromise between these issues is to instead rely on factor scores saved from preliminary measurement models (e.g., Kam, Morin, Meyer, & Topolnytsky, 2015; Morin & Marsh, 2015). Although factor scores do not explicitly control for measurement errors with equal efficacy as latent variables, they can provide a partial implicit control for measurement errors by giving differentially more weight to the more reliable items (i.e., those items with stronger factor loadings). As such, they provide researchers with a reasonable alternative to the use of observed scale scores. In addition, factor scores tend to preserve the nature of the underlying measurement structure in a superior manner compared to scale scores (e.g., can provide methodological controls, account for a bifactor solution or cross-loadings; Morin, Boudrias et al., 2016, 2017). Notably, the measurement structure of a multidimensional instrument may not be as clear as a set of observed scale scores would reflect. More specifically, the simple structure of a multidimensional inventory should be verified prior to using the focal variables in a mixture model. This verification may potentially lead to investigating the psychometric structure of an inventory with exploratory structural equation modeling and/or bifactor exploratory structural equation modeling (Marsh, Morin, Parker, & Kaur, 2014; Marsh, Nagengast, & Morin, 2013; McLarnon & Tarraf, 2017; Taku & McLarnon, 2018). Without accounting for an inventory's complex measurement structure, a resulting

profile solution may not accurately represent the underlying phenomenon and classification of interest, and may therefore have questionable construct validity. A third advantage of factor scores, which might tend to be crucially important in cross-cultural and/or longitudinal research, is that preliminary measurement models can be used to systematically assess the invariance of a measure across groups and/or time points (Millsap, 2011). In these research designs, factor scores can be saved from the most invariant measurement model, ensuring comparability across groups and/or time points. A final advantage of factors scores comes from their natural standardization, making them directly and easily interpretable with a mean of zero and a variance of one.

RANDOM STARTS

Mixture models are estimated through an iterative process that risks converging on a local solution, rather than on a true maximum likelihood, when the number of starting values is not sufficient (Hipp & Bauer, 2006). Therefore, best practice is to estimate mixture models using many sets of random starting values, and allowing each set of starting values to iterate sufficiently (Hipp & Bauer, 2006; McLachlan & Peel, 2000). Despite a general recognition that “the more, the better” applies to sets of random starting values, increasing the number of random starts also increases the computation time for each model under investigation. In practice, our recommendation is to use at least 3000 sets of random starting values, allowing up to 100 iterations for each of these sets, and retaining the 100 best sets of starting values (i.e., those that have resulted in the highest loglikelihood estimates) for a final stage of optimizations. The *Mplus* syntax requesting these specifications is given below, as these recommendations override the *Mplus* defaults. These values can be increased to 5000, 200, and 200, respectively, when the final solution is not sufficiently replicated. Moreover, we see these values as minimums that can be increased as needed. We often use even larger values in our own research (e.g., 10000, 1000, 500, respectively).

For interested readers, Masyn (2013) presented an introduction to maximum likelihood estimation and the use of an expectation-maximization algorithm in mixture modeling.

CLASS ENUMERATION

Alternative profile solutions that extract differing numbers of latent profiles (often ranging from one profile to the number of profiles that exceeds theoretical expectations, though rarely greater than 10 profiles) are typically contrasted in order to select the final, optimal solution. Selection of the optimal model is determined on the basis of considering (a) the substantive meaning and theoretical conformity of the solution, (b) the statistical adequacy of the solution, and (c) the statistical indicators and model fit indices accompanying each solution.

Focusing on the statistical indicators of model-data fit: the Akaike Information Criterion (AIC), the Consistent AIC (CAIC), the Bayesian information criterion (BIC), and the sample-adjusted BIC (SABIC) are often reported as components of a comprehensive assessment of a profile solution's fit. Although the CAIC is not currently provided by *Mplus*, it can be easily calculated as $BIC + fp$, where fp is the number of free parameters. These information criteria are considered in terms of "lower is better," such that lower values indicate a better-fitting model. Traditional likelihood ratio tests (LRT) are inappropriate for profile enumeration purposes to compare, for instance, the loglikelihood values across models with two versus three profiles (LRTs can be appropriate for the comparisons of nested models based on the same variables and number of profiles, as in several of the models presented later in this supplemental material). For profile enumeration, alternative LRT approximations are available: the Lo, Mendel and Rubin (2001) LRT (the standard [LMR] and adjusted version of this test [ALMR], which typically yield the same conclusions), and the bootstrapped LRT (BLRT; McLachlan & Peel, 2000). These LRT variations compare a k -profile model with a $k-1$ -profile model. BLRT p -values $> .05$ indicate that the $k-1$ profile model should be retained over a k -profile model (see Nylund, Asparouhov, & Muthén, 2007). Finally, although it should not be

used for enumeration purposes, the entropy estimate associated with each model provides a summary of classification accuracy, with larger values (closer to 1) indicating greater accuracy.

Multiple simulation studies (for a recent review, see Diallo, Morin, & Lu, 2016) indicate that four of these indicators (the CAIC, BIC, ABIC, and BLRT) are particularly effective, while less emphasis should be placed on the AIC and LMR/ALMR (Diallo, Morin, & Lu, 2016, 2017; Henson et al., 2007; McLachlan & Peel, 2000; Nylund et al., 2007; Peugh & Fan, 2013, 2014; Tein, Coxe, & Cham, 2013; Tofighi & Enders, 2008; Tolvanen, 2007; Yang, 2006). When these indicators fail to retain the optimal model by indicating which model reaches a minimum on the information criteria, the BIC and CAIC tend to underestimate the true number of profiles, and the AIC, ABIC, and BLRT tend to overestimate it. The simulation study of Diallo, Morin, and Lu (2017) further suggested that the BIC and CAIC should be privileged under conditions of high entropy ($\geq .80$), whereas the ABIC and BLRT perform better in conditions of low entropy ($\leq .60$). However, all of these indicators can also be heavily influenced by sample size (Marsh, Lüdtke, Trautwein, & Morin, 2009), which means that with sufficiently large samples, they may suggest a need for additional of profiles to reach adequate fit without ever converging on a solution that demonstrates the lowest values on these indices. When this happens, it is recommended to rely on an “elbow plot” (Morin, Maïano, et al., 2011; Petras & Masyn, 2010). These plots illustrate the gains in fit (i.e., decreasing values of the information criteria) related to each additional profile. The point after which the slope flattens suggests the optimal number of profiles. Examples of using elbow plots to inform judgments of profile enumeration can be found in Chen, Morin, Parker, and Marsh (2015), Ciarrochi, Morin, Sahdra, Litalien, and Parker (2017), Litalien, Morin, and McInerney (2017), McLarnon and O’Neill (2018), McLarnon et al. (2019), and Meyer, Morin, Stanley, and Maltin (2019), among others.

REFERENCES

- Asparouhov, T., & Muthén, B.O. (2012). Using *Mplus* TECH11 and TECH14 to test the number of latent classes (*Mplus* Web Note #14). Retrieved from <http://statmodel.com/examples/webnotes/webnote14.pdf>
- Asparouhov, T. & Muthén, B.O. (2014). Auxiliary variables in mixture modeling: Three-step approaches using *Mplus*. *Structural Equation Modeling*, *21*, 329–341. doi:10.1080/10705511.2014.915181
- Asparouhov, T., & Muthén, B.O. (2015). Auxiliary variables in mixture modeling: Using the BCH method in *Mplus* to estimate a distal outcome model and an arbitrary secondary model (*Mplus* Web Note #21). Retrieved from https://www.statmodel.com/download/asparouhov_muthen_2014.pdf
- Bakk, Z., & Kuha, J. (2018). Two-step estimation of models between latent classes and external variables. *Psychometrika*, *83*, 871-892. doi:10.1007/s11336-017-9592-7
- Bakk, Z., & Vermunt, J.K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling*, *23*, 20–31. doi:10.1080/10705511.2014.955104
- Berlin, K.S., Williams, N.A., & Parra, G.R. (2014). An introduction to latent variable mixture modeling (Part 1): Cross sectional latent class and latent profile analyses. *Journal of Pediatric Psychology*, *39*, 174–187. doi:10.1093/jpepsy/jst084
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Chénard Poirier, L.-A., Morin, A.J.S., & Boudrias, J.-S. (2017). On the merits of coherent leadership empowerment behaviors: A mixture regression approach. *Journal of Vocational Behavior*, *103*, 66–75. doi:10.1016/j.jvb.2017.08.003
- Chen, X., Morin, A.J.S., Parker, P.D., & Marsh, H.W. (2015). Developmental investigation of the domain-specific nature of the life satisfaction construct across the post-school

- transition. *Developmental Psychology*, *51*, 1074–1085. doi:10.1037/a0039477
- Ciarrochi, J., Morin, A.J.S., Sahdra, B.K., Litalien, D., & Parker, P.D. (2017). A longitudinal person-centered perspective on youth social support: Relations with psychological wellbeing. *Developmental Psychology*, *53*, 1154–1169. doi:10.1037/dev0000315
- Diallo, T.M.O., Morin, A.J.S., & Lu, H. (2016). Impact of misspecifications of the latent variance–covariance and residual matrices on the class enumeration accuracy of growth mixture models. *Structural Equation Modeling*, *23*, 507–531. doi:10.1080/10705511.2016.1169188
- Diallo, T.M.O., Morin, A.J.S., & Lu, H. (2017). The impact of total and partial inclusion or exclusion of active and inactive time invariant covariates in growth mixture models. *Psychological Methods*, *22*, 166–190. doi:10.1037/met0000084
- Eid, M., Langeheine, R., & Diener, E. (2003). Comparing typological structures across cultures by multigroup latent class analysis: A primer. *Journal of Cross-Cultural Psychology*, *34*, 195–210. doi:10.1177/0022022102250427
- Geiser, C., Lehmann, W., & Eid, M. (2006). Separating “rotators” from “nonrotators” in the Mental Rotations Test: A multigroup latent class analysis. *Multivariate Behavioral Research*, *41*, 261–293. doi:10.1207/s15327906mbr4103_2
- Henson, J.M., Reise, S.P., & Kim, K.H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling*, *14*, 202–226. doi:10.1080/10705510709336744
- Hipp, J.R., & Bauer, D.J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, *11*, 36–53. doi:10.1037/1082-989X.11.1.36
- Janssen, J.H.M., van Laar, S., de Rooij, M.J., Kuha, J., & Bakk, Z. (in press). The detection and modeling of direct effects in latent class analysis. *Structural Equation Modeling*.

doi:10.1080/10705511.2018.1541745

Kam, C., Morin, A.J.S., Meyer, J.P., & Topolnytsky, L. (2016). Are commitment profiles stable and predictable? A latent transition analysis. *Journal of Management*, *42*, 1462–1490.

doi:10.1177/0149206313503010

Kankaraš, M., Moors, G., & Vermunt, J.K. (2011). Testing for measurement invariance with latent class analysis. In E. Davidov, P. Schmidt & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 359–384). New York, NY: Routledge.

Landis, R.S., Edwards, B.D., & Cortina, J.M. (2009). On the practice of allowing correlated residuals among indicators in structural equation models. In C.E. Lance & R.J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 37–59). New York: Routledge.

Lanza, S.T., Tan, X., & Bray, B.C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling*, *20*, 1–26.

doi:10.1080/10705511.2013.742377

Litalien, D., Morin, A.J.S., & McInerney, D.M. (2017). Achievement goal profiles among adolescent males and females. *Developmental Psychology*, *53*, 731–751.

doi:10.1037/dev0000288

Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767–778. doi:10.1093/biomet/88.3.767

Marsh, H.W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares, & J. McArdle (Eds.), *Contemporary psychometrics* (pp. 275–340). Hillsdale, NJ: Erlbaum.

Marsh, H.W., Lüdtke, O., Trautwein, U., & Morin, A.J.S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling*, *16*,

191–225. doi:10.1080/10705510902751010

- Marsh, H.W., Morin, A.J.S., Parker, P., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85–110. doi:10.1146/annurev-clinpsy-032813-153700
- Marsh, H.W., Nagengast, B., & Morin, A.J.S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology, 49*, 1194–1218. doi:10.1037/a0026913
- Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T.D. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. 2, pp. 551–611). New York: Oxford University Press.
- McLachlan, G.J., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.
- McLarnon, M.J.W., & O'Neill, T.A. (2018). Extensions of auxiliary variable approaches for the investigation of mediation, moderation, and conditional effects in mixture models. *Organizational Research Methods, 21*, 955–982. doi:10.1177/1094428118770731
- McLarnon, M.J.W., & Tarraf, R.C. (2017). The Dark Triad: Specific or general sources of variance? A bifactor exploratory structural equation modeling approach. *Personality and Individual Differences, 112*, 67–73. doi:10.1016/j.paid.2017.02.049
- McLarnon, M.J.W., Carswell, J.J., & Schneider, T.J. (2015). A case of mistaken identity? Latent profiles in vocational interests. *Journal of Career Assessment, 23*, 166–185. doi:10.1177/1069072714523251
- McLarnon, M.J.W., DeLongchamp, A.C., & Schneider, T.J. (2019). Faking it! Individual differences in types and degrees of faking behavior. *Personality and Individual Differences, 138*, 88–95. doi:10.1016/j.paid.2018.09.024
- McNeish, D., Stapleton, L.M., & Silverman, R.D. (2017). On the unnecessary ubiquity of

- hierarchical linear modeling. *Psychological Methods*, 22, 114–140.
doi:10.1037/met0000078
- Mehta, P.D., & West, S.G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, 5, 23–43. doi:10.1037/1082-989X.5.1.23
- Meyer, J.P., Morin, A.J.S., Stanley, L.J., & Maltin, E.R. (2019). Teachers' dual commitment to the organization and occupation: A person-centered investigation. *Teaching and Teacher Education*, 77, 100–111. doi:10.1016/j.tate.2018.09.009
- Millsap, R.E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
- Morin, A.J.S., & Litalien, D. (2017). *Webnote: Longitudinal tests of profile similarity and latent transition analyses*. Retrieved January 5, 2019 from https://smslabstats.weebly.com/uploads/1/0/0/6/100647486/lt_a_distributional_similarity_v0_2.pdf. Montreal, QC: Substantive Methodological Synergy Research Laboratory.
- Morin, A.J.S., & Marsh, H.W. (2015). Disentangling shape from level effects in person-centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling*, 22, 39–59.
doi:10.1080/10705511.2014.919825
- Morin, A.J.S., Boudrias, J.-S., Marsh, H.W., Madore, I., & Desrumaux, P. (2016). Further reflections on disentangling shape and level effects in person-centered analyses: An illustration aimed at exploring the dimensionality of psychological health. *Structural Equation Modeling*, 23, 438–454. doi:10.1080/10705511.2015.1116077
- Morin, A.J.S., Boudrias, J.-S., Marsh, H.W., McInerney, D.M., Dagenais-Desmarais, V., Madore, I., & Litalien, D. (2017). Complementary variable- and person-centered approaches to exploring the dimensionality of psychometric constructs: Application to psychological wellbeing at work. *Journal of Business and Psychology*, 32, 395–419.
doi:10.1007/s10869-016-9448-7

- Morin, A.J.S., Maïano, C., Nagengast, B., Marsh, H.W., Morizot, J., & Janosz, M. (2011). General growth mixture analysis of adolescents' developmental trajectories of anxiety: The impact of untested invariance assumptions on substantive interpretations. *Structural Equation Modeling, 18*, 613–648. doi:10.1080/10705511.2011.607714
- Morin, A.J.S., Meyer, J.P., Creusier, J., & Biétry, F. (2016). Multiple-group analysis of similarity in latent profile solutions. *Organizational Research Methods, 19*, 231–254. doi:10.1177/1094428115621148
- Morin, A.J.S., Morizot, J., Boudrias, J.-S., & Madore, I. (2011). A multifoci person-centered perspective on workplace affective commitment: A latent profile/factor mixture analysis. *Organizational Research Methods, 14*, 58–90. doi:10.1177/1094428109356476
- Morin, A.J.S., Scalas, L.F., & Marsh, H.W. (2015). Tracking the elusive actual-ideal discrepancy model within latent subpopulations. *Journal of Individual Differences, 36*, 65–72. doi:10.1027/1614-0001/a000157
- Muthén, B.O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 267–316. doi:10.2307/271070
- Muthén, L.K., & Muthén, B.O. (2017). *Mplus User's Guide* (8th ed.). Los Angeles: Muthén & Muthén.
- Nylund, K.L., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535–569. doi:10.1080/10705510701575396
- Nylund-Gibson, K.L., Grimm, R., Quirk, M., & Furlong, M. (2014). A latent transition mixture model using the three-step specification. *Structural Equation Modeling, 21*, 439–454. doi:10.1080/10705511.2014.915375
- O'Neill, T.A., McLarnon, M.J.W., Hoffart, G.C., Woodley, H.J., & Allen, N.J. (2018). The

- structure and function of team conflict state profiles. *Journal of Management*, *44*, 811–836. doi:10.1177/0149206315581662
- O’Neill, T.A., McLarnon, M.J.W., Xiu, L., & Law, S.J. (2016). Core self-evaluations, perceptions of group potency, and job performance: The moderating role of individualism and collectivism cultural profiles. *Journal of Occupational and Organizational Psychology*, *89*, 447–473. doi:10.1111/joop.12135
- Pastor, D.A., Barron, K.E., Miller, B.J., & Davis, S.L. (2007). A latent profile analysis of college students’ achievement goal orientation. *Contemporary Educational Psychology*, *32*, 8–47. doi:10.1016/j.cedpsych.2006.10.003
- Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A.R. Piquero, & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 69–100). New York: Springer.
- Peugh, J., & Fan, X. (2013). Modeling unobserved heterogeneity using latent profile analysis: A Monte Carlo simulation. *Structural Equation Modeling*, *20*, 616–639. doi:10.1080/10705511.2013.824780
- Peugh, J., & Fan, X. (2014). Enumeration index performance in generalized growth mixture models: A Monte Carlo test of Muthén’s (2003) hypothesis. *Structural Equation Modeling*, *23*, 115–131. doi:10.1080/10705511.2014.919823
- Raykov, T., & Marcoulides, G.A. (2004). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling*, *11*, 621–637. doi:10.1207/s15328007sem1104_7
- Taku, K., & McLarnon, M.J.W. (2018). Posttraumatic growth profiles and their relationships with HEXACO personality traits. *Personality and Individual Differences*, *134*, 33–42. doi:10.1016/j.paid.2018.05.038
- Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of

- classes in latent profile analysis. *Structural Equation Modeling*, 20, 640–657.
doi:10.1080/10705511.2013.824781
- Tofighi, D., & Enders, C. (2008). Identifying the correct number of classes in growth mixture models. In G.R. Hancock & K.M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 317–341). Charlotte, NC: Information Age.
- Tolvanen, A. (2007). *Latent growth mixture modeling: A simulation study*. PhD dissertation, Department of Mathematics, University of Jyväskylä, Jyväskylä, Finland.
- Vermunt, J.K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18, 450–469. doi:10.1093/pan/mpq025
- Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. West Sussex, UK: John Wiley & Sons.
- Yang, C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Computational Statistics & Data Analysis*, 50, 1090–1104. doi:10.1016/j.csda.2004.11.004

Appendix B

Mplus Syntax for the Estimation of Mixture Models

In *Mplus* syntax, annotations can be preceded by an exclamation mark, !, and are not interpreted by the program. All *Mplus* commands are not case sensitive.

DATA

The first part of the *Mplus* syntax identifies the dataset to be used in the analysis. If the dataset is in the same folder as the input file (the recommended strategy), only the filename of the dataset needs to be indicated. If the dataset is in another folder then the full path needs to be specified. Here, the dataset, saved as a comma-separated values file (*.csv; but we also frequently use tab-delimited *.dat files) that can be opened and manipulated in Microsoft Excel or other standard spreadsheet software, is labeled “filename.csv” and is located in the same folder as the *Mplus* input file (i.e., the *.inp file containing the syntax).

TITLE: Mixture Modeling in Organizational Behavior Research

!A title is not a required syntax component, but can be helpful to track files and results

DATA:

FILE IS “filename.csv”;

VARIABLE

In the VARIABLE section of the syntax, the NAMES function identifies all variables included in the dataset, in order of appearance. USEVARIABLES specifies the variables to be used in the analysis. Here, only the variables IND1_T1, IND2_T1, and IND3_T1 are being considered as indicators of a latent profile (IND is used as a generic label for “indicator”; 1, 2, and 3 represents the *i* set of indicators; and _T1 reflects Time 1 measures). Variable names should be restricted to eight characters or less. MISSING defines the values in each variable that denote a missing piece of data (typically the same value is used for all variables). Here, any instance of

“999” would be interpreted by *Mplus* as a missing value. The `IDVARIABLE` function defines the unique identifier for each case (i.e., participant, employee, team, etc.). `CLASSES` defines the number of latent profiles estimated (here, 3), with the “C” preceding the number in parentheses being an arbitrary label assigned to the latent categorical variable that defines the latent profiles to be extracted. As noted, most typically, multiple LPAs are conducted, in a somewhat exploratory approach, by varying the number of profiles extracted, so in other analyses models with 1, 2, 4, 5, etc. profiles would also be specified.

`USEOBSERVATIONS` can be used to limit the estimation to a subset of cases. In this example, if the `!` is removed, the function would limit the estimation to women (identified by the `Gender` variable, which might have the coding of 1 = males; 2 = females) and would exclude men and cases with missing gender information. The `CLUSTER` function can be used to define a unique identifier for a clustering variable (i.e., a level-2 variable) to be controlled for or accounted for in the analysis (e.g., teams, classrooms, etc.), or to enable multilevel specifications. By default, *Mplus* considers variables as continuous. However, *Mplus* can accommodate categorical (binary or polytomous), censored, count, or nominal variables (see Muthén & Muthén, 2017). The appropriate variables would just be specified by each of these respective commands (and the exclamation point removed) to be incorporated into an analysis. Notably, the `NOMINAL` specification will be used later in this material to illustrate the manual implementation of the three-step procedure used to investigate covariate relations.

```
VARIABLE:
```

```
NAMES = ID Gender Team_ID Pred1 Pred2 Cor1 Cor2 Out1 Out2
```

```
IND1_T1 IND2_T1 IND3_T1
```

```
IND1_T2 IND2_T2 IND3_T2;
```

```
USEVARIABLES = IND1_T1 IND2_T1 IND3_T1;
```

```
! CATEGORICAL = ;
```

```
! CENSORED = ;  
! COUNT = ;  
! NOMINAL = ;  
MISSING = all (999);  
IDVARIABLE = ID;  
CLASSES = C(3);  
CLUSTER = Team_ID;  
! USEOBSERVATIONS Gender EQ 2;
```

ANALYSIS

The ANALYSIS section describes the technical specifications of the LPA itself. Here, we request the estimation of a mixture model (TYPE = MIXTURE;), which includes a correction for the nesting of cases within a higher-level unit (i.e., work team; TYPE = COMPLEX;). Further, we request that *Mplus* uses its robust maximum likelihood estimator (ESTIMATOR = MLR;), which provides standard errors that are robust to non-normality, and to non-independence when used in conjunction with TYPE=COMPLEX; (McNeish, Stapleton, & Silverman, 2017; Muthén & Satorra, 1995; Muthén & Muthén, 2017; *Mplus* uses MLR by default for mixture models so it is not necessary to declare this explicitly). STARTS = 3000 100; requests 3000 initial sets of random start values, with the best 100 of these starts (i.e., those that result in the highest loglikelihood values; see below) being retained for a final set of optimization iterations. STITERATIONS = 100; requests that all random starts be allowed a maximum of 100 iterations. PROCESS = 3; requests that 3 of the available processors in the local computer be used to speed up the estimation (i.e., 1/3 of the random starts will be allocated to each processor to distribute the model estimation and complete it in parallel).

ANALYSIS:

TYPE = MIXTURE COMPLEX;

ESTIMATOR = MLR;

PROCESS = 3;

STARTS = 3000 100;

STITERATIONS = 100;

As part of the output, *Mplus* provides the loglikelihood values associated with each of the random starting value runs that were retained for the final stage of optimization iterations and reached convergence. The output will also indicate how many of the start value runs did not converge (i.e., the “perturbed” values; in this example there are 15 [of 100] that did not reach convergence). The first number in this list gives the loglikelihood value (−6095.887), the second number (991399) is the random seed value that resulted in this loglikelihood, and the third number (1433) is the ordering within the 3000 random starting value sets that resulted in this random seed and associated loglikelihood.

RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST

LOGLIKELIHOOD VALUES

15 perturbed starting value run(s) did not converge.

Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:

−6095.887	991399	1433
−6095.887	165268	2436
−6095.887	551639	55
−6095.887	58353	1723
−6095.887	168648	1788
−6095.887	445692	2796

-6095.887	973369	202
-6095.887	623887	2611
-6096.828	163110	584
-6096.828	762461	425
-6096.828	62715	2599
-6104.647	264951	1130
-6104.647	930872	277
-6104.647	670998	1876
-6104.698	98068	998
-6105.491	361131	2563
...		

In this example, the best loglikelihood value was replicated eight times (in bold; the number of times the value of -6095.887 appears in the first column), which is satisfactory. Although no clear-cut rule exists, we suggest that solutions should be replicated at least three times.

Otherwise, the results may not be robust, and these loglikelihood values could reflect a local solution, rather than a global optimum. Should a model fail to sufficiently replicate the best loglikelihood, additional tests should be conducted with an increased number of random starting value sets and/or iterations (e.g., `STARTS = 10000 1000`; `STITERATIONS = 1000`); or by using user-defined start values. Alternatively, the values from the best-fitting, non-replicated solution, which are provided when `SVALUES` is requested in the `OUTPUT` section of the syntax (see below), can be copied into the `MODEL` portion of the syntax while keeping the random starts function active (we provide an example below). As noted, the second column provides the random seed associated with each random starting value set. Using the seed number can provide an easy way to replicate the final solution (or any other solution) while

decreasing computational time (i.e., the same loglikelihood value and parameter estimates will result). To use this procedure, the following ANALYSIS section has been adapted to replicate the above solution by inputting the seed number into the OPTSEED command. The usage of the OPTSEED however will not ensure that the solution is replicated if additional variables (i.e., covariates, predictors, or outcomes) are directly incorporated into the model. We address this issue in greater detail below.

```
ANALYSIS:  
  
TYPE = MIXTURE COMPLEX;  
  
ESTIMATOR = MLR;  
  
PROCESS = 3;  
  
STARTS = 0;  
  
OPTSEED = 991399;  
  
STITERATIONS = 100;
```

MODEL

The MODEL section of the syntax describes the specific structure of the mixture model to be estimated. We provide code, in sequence, for each of the models described in the chapter.

OUTPUT

This section of the syntax covers specific output provided by *Mplus*. Here, we request standardized model parameters (STDYX), sample statistics (SAMPSTAT), confidence intervals (CINTERVAL), the final estimates provided by the model that could be used as starting values for a subsequent analyses (SVALUES), the model residuals (RESIDUAL), the arrays of parameters (TECH1), the profile-specific descriptives (TECH7), the LMR and ALMR tests (TECH11), and the BLRT (TECH14). Since the BLRT can be computationally intensive, users may wish to omit it from an initial set of analyses, and then revisit each solution following the

procedure outlined by Asparouhov and Muthén (2012) to save time with complex models.

OUTPUT:

```
STDYX SAMPSTAT CINTERVAL SVALUES RESIDUAL TECH1 TECH7 TECH11
```

```
TECH14;
```

LATENT PROFILE ANALYSIS

The MODEL section includes an %OVERALL% portion that describes the global relations estimated among the variables that will apply to each profile. Profile-specific statements, %C#1%, %C#2%, and %C#3% (where C is the arbitrary label used to define the latent categorical variable from the CLASSES command of the VARIABLE section, and the number [1 to k] refers to the specific nominal category of this variable [i.e., the specific profile]). Here, as no commands are presented directly beneath the %OVERALL% section, no relations between variables will be constrained to equality across profiles (though see later examples and Pastor, Barron, Miller, & Davis, 2007). As noted, throughout our examples, three profiles are estimated from three indicators taken at the first time point (IND1_T1, IND2_T1, and IND3_T1). The profile specific sections request that the means (indicated by the name of the variable within square brackets []) and variances (defined by the names only) of the indicators be freely estimated in all profiles.

MODEL:

```
%OVERALL%
```

```
%C#1%
```

```
[IND1_T1 IND2_T1 IND3_T1];
```

```
IND1_T1 IND2_T1 IND3_T1;
```

```

%C#2%

[IND1_T1 IND2_T1 IND3_T1];

IND1_T1 IND2_T1 IND3_T1;

%C#3%

[IND1_T1 IND2_T1 IND3_T1];

IND1_T1 IND2_T1 IND3_T1;

```

Profiles can also be estimated with the variances of respective indicators held equal across profiles by removing, or deactivating with !, the profile-specific references of each variable's variance:

```

MODEL:

%OVERALL%

%C#1%

[IND1_T1 IND2_T1 IND3_T1];

%C#2%

[IND1_T1 IND2_T1 IND3_T1];

%C#3%

[IND1_T1 IND2_T1 IND3_T1];

```

In *Mplus*, correlations are identified using a *WITH* command, so should a researcher wish to allow correlated uniquenesses among profile indicators, where the correlations are constrained to equality across profiles, the following could be added to the *%OVERALL%* section of the

syntax. However, we generally would not recommend this strategy, and would encourage researchers to use this approach with caution and only with a strong theoretical rationale, given the well-known issues associated with correlated uniquenesses, which are known to drastically change the meaning of the estimated latent variables (here the profiles; see Landis, Edwards, & Cortina, 2009).

```

MODEL:

%OVERALL%

IND1_T1 WITH IND2_T1 IND3_T1;
IND2_T1 WITH IND3_T1;

%C#1%

[IND1_T1 IND2_T1 IND3_T1];
IND1_T1 IND2_T1 IND3_T1;

%C#2%

[IND1_T1 IND2_T1 IND3_T1];
IND1_T1 IND2_T1 IND3_T1;

%C#3%

[IND1_T1 IND2_T1 IND3_T1];
IND1_T1 IND2_T1 IND3_T1;

```

To estimate a model with correlated uniquenesses among all profile indicators, which are freely estimated in all profiles rather than constrained to equality, the following commands can be adapted. As above, we would not generally recommend this strategy.

MODEL:

%OVERALL%

IND1_T1 WITH IND2_T1 IND3_T1;

IND2_T1 WITH IND3_T1;

! Note, even though the general commands appear here,

! they are “over-ridden” with the profile-specific commands below

%C#1%

[IND1_T1 IND2_T1 IND3_T1];

IND1_T1 IND2_T1 IND3_T1;

IND1_T1 WITH IND2_T1 IND3_T1;

IND2_T1 WITH IND3_T1;

%C#2%

[IND1_T1 IND2_T1 IND3_T1];

IND1_T1 IND2_T1 IND3_T1;

IND1_T1 WITH IND2_T1 IND3_T1;

IND2_T1 WITH IND3_T1;

%C#3%

[IND1_T1 IND2_T1 IND3_T1];

IND1_T1 IND2_T1 IND3_T1;

```
IND1_T1 WITH IND2_T1 IND3_T1;
IND2_T1 WITH IND3_T1;
```

When SVALUES are requested in the output, *Mplus* will provide a section, similar to the following. This section can be copied and pasted and used as a replacement for the MODEL section. In conjunction with setting the STARTS command to 0 (STARTS = 0) the final solution will be replicated without the computational requirements of the random starts. This function may be particularly useful when including covariates, and may minimize the occurrence of changes in an LPA following the inclusion of covariates to the model. However, if there is substantial change (e.g., proportions of cases assigned to each profile, means of indicators in each profile) between a final unconditional model and a conditional model with covariates, the researcher should use one of the procedures described below (i.e., the three-step, DCAT, or BCH approaches; see McLarnon & O'Neill, 2018). Below, we provide an overview of McLarnon and O'Neill's (2018) manual approach that can be implemented to incorporate auxiliary variables when a shift in a profile solution has occurred following the implementation of direct inclusion approaches.

```
MODEL: ! COMMANDS WITH FINAL ESTIMATES USED AS STARTING VALUES

%OVERALL%

[C#1*-0.93515];

[C#2*0.49113];

%C#1%

[IND1_T1*-1.59367];

[IND2_T1*-0.93155];

[IND3_T1*0.39664];
```

```

IND1_T1*0.53789;
IND2_T1*2.12542;
IND3_T1*1.59137;

%C#2%
[IND1_T1*-0.18248];
[IND2_T1*-0.29812];
[IND3_T1*-0.15201];
IND1_T1*0.42376;
IND2_T1*0.43116;
IND3_T1*0.72513;

%C#3%
[IND1_T1*0.94111];
[IND2_T1*0.86526];
[IND3_T1*0.08333];
IND1_T1*0.44053;
IND2_T1*0.48030;
IND3_T1*1.13489;

```

LATENT PROFILE ANALYSIS WITH COVARIATES

Predictors

Direct inclusion of predictors into the model is done by way of multinomial logistic regressions where the predictors are specified as having an impact on profile membership (i.e., C#1-C#2 ON Pred1 Pred2;), where one fewer C# statement than the total number of profiles is necessary

in the %OVERALL% section. The C# portions of the command reflect the multinomial probability of membership in the various categories of the latent profile variable, which represent the relative size of each profile and the proportion of the sample assigned to each profile.

```

MODEL:
%OVERALL%
C#1-C#2 ON Pred1 Pred2;

%C#1%
[IND1_T1 IND2_T1 IND3_T1];
IND1_T1 IND2_T1 IND3_T1;

%C#2%
[IND1_T1 IND2_T1 IND3_T1];
IND1_T1 IND2_T1 IND3_T1;

%C#3%
[IND1_T1 IND2_T1 IND3_T1];
IND1_T1 IND2_T1 IND3_T1;

```

To enhance the probability that the nature of the profiles will remain unchanged by the inclusion of predictors, the SVALUES from the final solution can be used (as noted previously), and the random starts function should be deactivated (STARTS = 0;). If this is not sufficient to prevent shifts in the model parameters, the class specific parameters (i.e., variable means and variances within the %C#1% to %C#3% profile-specific sections) can be fixed

(using @ rather than *; see alternative codes below; see also Bakk & Kuha, 2018; Janssen, van Laar, de Rooij, Kuha, & Bakk, in press). The start values in the %OVERALL% section to describe that the class probabilities should not be fixed (i.e., leave these with an asterisk; [C#1*-0.93515]:), as the inclusion of predictors changes the meaning of these parameter from reflecting class probabilities to reflecting the multinomial intercepts of these class probabilities as predicted by the covariates.

```
%OVERALL%
```

```
[C#1*-0.93515]; [C#2*0.49113];
```

```
C#1-C#2 ON Pred1 Pred2;
```

```
%C#1%
```

```
[IND1_T1*-1.59367]; ! Alternatively: [IND1_T1@-1.59367];
```

```
[IND2_T1*-0.93155]; ! Alternatively: [IND2_T1@-0.93155];
```

```
[IND3_T1*0.39664]; ! Alternatively: [IND3_T1@0.39664];
```

```
IND1_T1*0.53789;
```

```
IND2_T1*2.12542;
```

```
IND3_T1*1.59137;
```

```
%C#2%
```

```
[IND1_T1*-0.18248];
```

```
IND2_T1*-0.29812];
```

```
[IND3_T1*-0.15201];
```

```
IND1_T1*0.42376;
```

```
IND2_T1*0.43116;
```

```
IND3_T1*0.72513;
```

```

%C#3%
[IND1_T1*0.941111];
[IND2_T1*0.86526];
[IND3_T1*0.08333];
IND1_T1*0.44053;
IND2_T1*0.48030;
IND3_T1*1.13489;

```

In contrast to the previous approach in which predictors are directly included into the model, *Mplus* also includes a number of approaches that facilitate examining predictor, covariate, and outcome relations while maintaining these variables as external, or auxiliary, to the focal mixture model. Advantageously, these approaches preserve the construct validity of an unconditional profile solution by reducing the likelihood of a shift in the structure and definition of a profile. In particular, with the direct inclusion approach offered in the previous syntax, there is a potential for confounding in the meaning and nature of the latent profiles because covariates can directly contribute information into the estimation of the mixture model. This means that the profiles could come to reflect the heterogeneity not only in the focal profile indicators but also in the covariates, potentially raising concerns over the nature of the construct assessed by the profiles. Accordingly, direct inclusion raises the possibility that a latent profile “could lose its meaning” (Asparouhov & Muthén, 2014, p. 329). To provide an alternative to the direct inclusion approach when it results in substantial shifts in the nature of an optimal unconditional profile solution, *Mplus* includes a number of *auxiliary* procedures for examining relations between a latent profile and covariates without allowing them to influence the nature of the profiles. In these contexts, covariates are referred to as *auxiliary* variables. The manual implementation of these procedures forms the basis of McLarnon and O’Neill’s

(2018) framework for examining auxiliary covariate relations (which is based on Asparouhov & Muthén, 2014, 2015; Nylund-Gibson, Grimm, Quirk, & Furlong, 2014).

Among the available AUXILIARY approaches automated in *Mplus*, the R3STEP procedure (see Asparouhov & Muthén, 2014; Vermunt, 2010) is currently optimal for examining predictor relations. This approach is similar to the multinomial logistic regressions described above, but explicitly tests, through an automated procedure, whether including the predictors result in a change in the nature of the profiles. When this occurs (and the previous direct inclusion approaches do not work), then predictors should be treated as correlates (see below). To use R3STEP, the following line of code needs to be included to the VARIABLE section (in bold). The MODEL section remains unchanged (which can be specified with or without starting values).

```
VARIABLE:
NAMES = ID Gender Team_ID Pred1 Pred2 Cor1 Cor2 Out1 Out2
IND1_T1 IND2_T1 IND3_T1
IND1_T2 IND2_T2 IND3_T2;
USEVARIABLES = IND1_T1 IND2_T1 IND3_T1;
MISSING = all (999);
IDVARIABLE = ID;
CLASSES = C(3);
AUXILIARY = Pred1 (R3STEP) Pred2 (R3STEP);
```

Outcomes

The direct inclusion of outcomes involves adding them as additional mixture indicators, preferably while using the SVALUES from the final solution to provide a measure of stability in the nature of the profiles. Parameter labels are given to the means of the outcome variables

in each profile (in parentheses), and these labels can subsequently be used within a MODEL CONSTRAINT command to provide tests of significant differences (based on the delta method; Raykov & Marcoulides, 2004) of the outcome means between profiles. A MODEL TEST command can also be enabled to provide a Wald χ^2 test, as is provided by the automated three-step and BCH functions.

```
%OVERALL%
[C#1*-0.93515]; [C#2*0.49113];

%C#1%
[IND1_T1*-1.59367]; [IND2_T1*-0.93155]; [IND3_T1*0.39664];
IND1_T1*0.53789; IND2_T1*2.12542; IND3_T1*1.59137;
[Out1] (oa1);
[Out2] (ob1);
Out1;
Out 2;

%C#2%
[IND1_T1*-0.18248]; [IND2_T1*-0.29812]; [IND3_T1*-0.15201];
IND1_T1*0.42376; IND2_T1*0.43116; IND3_T1*0.72513;
[Out1] (oa2);
[Out2] (ob2);
Out1;
Out 2;

%C#3%
```

[IND1_T1*0.941111]; [IND2_T1*0.86526]; [IND3_T1*0.08333];

IND1_T1*0.44053; IND2_T1*0.48030; IND3_T1*1.13489;

[Out1] (oa3);

[Out2] (ob3);

Out1;

Out 2;

MODEL CONSTRAINT:

! New parameters are created using this function to reflect pairwise mean differences between

! profiles (e.g., y12 reflects the differences between the Out1 means in profiles 1 and 2)

NEW (y12);

y12 = oa1-oa2;

NEW (y13);

y13 = oa1-oa3;

NEW (y23);

y23 = oa2-oa3;

NEW (z12);

z12 = ob1-ob2;

NEW (z13);

z13 = ob1-ob3;

NEW (z23);

z23 = ob2-ob3;

MODEL TEST:

!A MODEL TEST provides an omnibus test of the specified null hypothesis

!(i.e., means in each profile are equivalent), so may only be meaningful if conducted

!once for each outcome. Thus, in a second run, the ! would be removed for the ob1-ob3 !labels, but then added to the oa1-oa3 labels. So that two, separate, Wald chi-square tests ! with 2 degrees of freedom are estimated.

oa1 = oa2;

oa2 = oa3;

!ob1 = ob2;

!ob2 = ob3;

Several AUXILIARY approaches (see Asparouhov & Muthén, 2014) are also available for outcome relations, which may be particularly useful when direct inclusion results in shifting profile parameters. The first is similar to the R3STEP approach and includes an assessment of the degree to which continuous outcomes change the nature of the profile solution. This approach can either allow the variances of the outcomes to be freely estimated across profiles (DU3STEP) or held equal across profiles (DE3STEP). A more recent alternative (BCH) has been shown to outperform these approaches, and ensures the stability of the profile solutions (Bakk & Vermunt, 2016). Finally, an alternative approach that has been shown to be advantageous in some situations can accommodate continuous (DCON) and categorical (DCAT) outcomes (Asparouhov & Muthén, 2015; Lanza, Tan, & Bray, 2013). As noted in the chapter, our current recommendations are to use the R3STEP procedure for predictors, the DCAT procedure for binary, categorical, and nominal outcomes, and the DU3STEP (the homogeneity of variance assumption underlying DE3STEP may not be realistic in many situations), DCON, or BCH functions for continuous outcomes (see Asparouhov & Muthén, 2014, 2015; Muthén & Muthén, 2017).

! Pick between these alternatives depending on nature of outcome variable:

AUXILIARY = Out1 (DU3STEP) Out2 (DU3STEP);

AUXILIARY = Out1 (DE3STEP) Out2 (DE3STEP);

AUXILIARY = Out1 (BCH) Out2 (BCH);

AUXILIARY = Out1 (DCON) Out2 (DCON);

AUXILIARY = Out1 (DCAT) Out2 (DCAT);

Correlates

Correlates can be incorporated to the model via the AUXILIARY (E) command, which does not assume directionality in the associations between profiles and covariates, and is thus well suited to correlates. We note however, that statistical research has shown this approach to be clearly suboptimal when considering outcome variables, for which it has been superseded by the BCH, DCON, DCAT, and DU3STEP/DE3STEP approaches (Asparouhov & Muthén, 2015). For this reason, and pending additional evidence, we urge caution with its use and suggest that researchers using the E command may want to verify the extent to which these results replicate when using the 3STEP, DCAT, DCON, or BCH methods, as appropriate.

AUXILIARY = Cor1 (E) Cor2 (E);

PROCEDURE FOR THE MANUAL IMPLEMENTATION OF AUXILIARY APPROACHES

As noted in the chapter, it is important to ensure that covariate inclusion does not result in the modification of the nature of the most optimal unconditional solution given that such a change would violate the theoretically driven assumption of the directionality of the associations between the covariates and the profiles (Marsh et al., 2009, Morin, Morizot et al., 2011). The various auxiliary approaches presented earlier were all developed to circumvent this problem

when it happens in the context of the direct inclusion approach. However, not all of the automated auxiliary approaches are immune to this problem (i.e., the three-step approaches remain sensitive to this issue), and none of them can be extended to tests of complex relations between profiles and covariates (e.g., mediation, moderation). To address this issue, particularly when bringing a latent profile into a larger statistical model with multiple auxiliary variables, where, for example, the profile is considered as a moderator of an $X \rightarrow Y$ relation, or when the profile variable is considered a mediator, or when the conditional effects on an outcome (i.e., mean differences after accounting for a control variable) are of interest, McLarnon and O’Neill (2018) proposed leveraging the manual implementation of the three-step and BCH procedures (see also Morin & Litalien, 2017). McLarnon and O’Neill (2018) noted that preference could be placed on the BCH approach, but suggested that users could compare results from both approaches as an index of robustness. Implementing both the three-step and BCH approaches involve three analytical stages. We first outline the procedure of the BCH approach.

BCH Approach

First, in the final/optimal unconditional model a series of new variables that contain the classification information is requested within the `SAVEDATA` command.

!Commands for BCH approach

```
SAVEDATA:
FILE IS Output_BCH.dat;
SAVE IS BCHWEIGHTS;
```

Second, the new variables that were saved in the “Output_BCH.dat” file, and reflect classification probabilities for each case, are then incorporated into the main dataset:

```

DATA:

FILE IS filename2.csv;

VARIABLE:

NAMES = ID Gender Team_ID Pred1 Pred2 Cor1 Cor2 Out1 Out2

IND1_T1 IND2_T1 IND3_T1

IND1_T2 IND2_T2 IND3_T2

BCHw1 BCHw2 BCHw3;

!Where the above BCH variables reflect the BCH weights for each case

USEVARIABLES = Out1 Pred1 BCHw1 BCHw2 BCHw3;

TRAINING ARE BCHw1 BCHw2 BCHw3 (BCH);

! These BCH weights are then incorporated using a special command declaring
! each weight as a TRAINING variable.

! Note. Mplus will not run an empty model with only the BCH variables in the TRAINING
! command, so to ensure that classification proportions with the BCH variables are correct as
! compared to Step 1 at least one other variable must be declared in USEVARIABLE

MISSING = all (999);

IDVARIABLE = ID;

CLASSES = C(3);

```

Third, the auxiliary model is specified. Here, we limit our demonstration to a straightforward example of the profiles moderating the relation between Pred1 and Out1. Readers can consult McLarnon & O'Neill (2018) for further details about the mediation and conditional effect parameterization.

MODEL:

%OVERALL%

Out1 ON Pred1;

%C#1%

[Out1] (m1_1);

Out1;

Out1 ON Pred1 (reg1);

%C#2%

[Out1] (m1_2);

Out1;

Out1 ON Pred1 (reg2);

%C#3%

[Out1] (m1_3);

Out1;

Out1 ON Pred1 (reg3);

MODEL TEST:

reg1 = reg2;

reg2 = reg3;

! Facilitates an omnibus chi-square test for the equivalence of regressions of Out1 on Pred1

! across profiles; analogous to F-change from typical moderated multiple regression

MODEL CONSTRAINT:

NEW(mod1_2 mod1_3 mod2_3);

```
mod1_2 = reg1-reg2;
```

```
mod1_3 = reg1-reg3;
```

```
mod2_3 = reg2-reg3;
```

! These constraints now compute pairwise comparisons of each profile's regression coefficients

Three-Step Approach

The first step of the three-step procedure utilizes a different SAVEDATA command from the final/optimal unconditional model.

!Commands for 3STEP approach

```
SAVEDATA:
```

```
FILE IS Output_VAM.dat;
```

```
SAVE IS CPROB;
```

The CPROB command associated with the three-step's SAVEDATA syntax saves the classification probabilities (which aren't used in the secondary analysis), and the modal profile variable, which reflects most likely profile membership for each participant and is used in subsequent analyses.

```
DATA:
```

```
FILE IS filename3.csv;
```

```
VARIABLE:
```

```
NAMES = ID Gender Team_ID Pred1 Pred2 Cor1 Cor2 Out1 Out2
```

```
IND1_T1 IND2_T1 IND3_T1
```

```
IND1_T2 IND2_T2 IND3_T2
```

```
CPROB1 CPROB2 CPROB3 MODAL;
```

! Where the CPROB variables reflect the membership probabilities of each case in each profile

USEVARIABLES = MODAL;

NOMINAL = MODAL;

MISSING = all (999);

IDVARIABLE = ID;

CLASSES = C(3);

CLUSTER = Team_ID;

*! In the MODEL command, the thresholds for nominal modal profile variable, MODAL, are
! then declared and fixed to values that account for membership uncertainty. These values
! are made available by Mplus from the Step 1 analyses (i.e., the final/optimal unconditional
! model), under a section of the output called: "Logits for the Classification Probabilities for
! the Most Likely Latent Class Membership (Column) by Latent Class (Row)"*

*! These values will differ across any different model from Step 1 and must be adjusted in this
! second step. This model with only MODAL in the USEVARIABLE statement can be run to
! check that profile membership is equal to the model from Step 1.*

MODEL:

%OVERALL%

%C#1%

[MODAL#1@0.994];

[MODAL#2@-1.581];

%C#2%

[MODAL#1@-2.575];

[MODAL#2@1.953];

%C#3%

[MODAL#1@-1.816];

[MODAL#2@-1.669];

Third, the auxiliary model is specified under the MODEL command. For sake of brevity, we re-illustrate the syntax from the BCH method where the profiles moderate the relation between Pred1 and Out1 (see McLarnon & O'Neill, 2018 for other modeling possibilities). See our description of the BCH procedure for an explanation of the MODEL TEST and MODEL CONSTRAINT commands.

```

MODEL:
%OVERALL%
Out1 ON Pred1;
%C#1%
[MODAL#1@0.994];
[MODAL#2@-1.581];
[Out1] (m1_1); Out1;
Out1 ON Pred1 (reg1);
%C#2%
[MODAL#1@-2.575];
[MODAL#2@1.953];
[Out1] (m1_2); Out1;
Out1 ON Pred1 (reg2);
%C#3%
[MODAL#1@-1.816];
[MODAL#2@-1.669];
[Out1] (m1_3); Out1;
Out1 ON Pred1 (reg3);
MODEL TEST:
reg1 = reg2;

```

```
reg2 = reg3;
```

```
MODEL CONSTRAINT:
```

```
NEW(mod1_2 mod1_3 mod2_3);
```

```
mod1_2 = reg1-reg2;
```

```
mod1_3 = reg1-reg3;
```

```
mod2_3 = reg2-reg3;
```

FACTOR MIXTURE ANALYSIS

Presented next is the input for the factor mixture analyses that can be used as a way to control for a global factor that is shared among the indicators (e.g., global levels of life satisfaction: Morin et al., 2017; global interest in general vocational activities: McLarnon et al., 2015).

Factor mixture analyses may facilitate estimating clearer latent profiles that are distinguished by shape and present different configurations, rather than profiles differentiated merely on the basis of different levels (see Morin & Marsh, 2015). The only difference with the previous LPA models is the introduction of a common factor model in the %OVERALL% section. This factor model is specified as invariant across profiles. This common factor is labeled G, and is defined by the same indicators that are used to estimate the latent profiles. In *Mplus*, BY defines factor loadings, and denotes use of IND1_T1, IND2_T1, and IND3_T1 as indicators of the factor, G. All loadings on this factor are freely estimated (the * associated with the first indicator overrides the default of constraining the loading of the first indicator to be 1.00, as in the referent indicator approach to identifying a factor model). Therefore, for identification purposes the factor variance is fixed to 1.00 (the @ is used to fix a parameter to a specific value). Because the intercepts of the indicators of this factor will be freely estimated across profiles, the factor mean must also be fixed to 0 in each profile for identification purposes.

MODEL:

%OVERALL%

G BY IND1_T1* IND2_T1 IND3_T1;

G@1;

[G@0];

%C#1%

[IND1_T1 IND2_T1 IND3_T1]; IND1_T1 IND2_T1 IND3_T1;

%C#2%

[IND1_T1 IND2_T1 IND3_T1]; IND1_T1 IND2_T1 IND3_T1;

%C#3%

[IND1_T1 IND2_T1 IND3_T1]; IND1_T1 IND2_T1 IND3_T1;

MULTIPLE-GROUP LATENT PROFILE ANALYSES

Configural Similarity

Returning to the VARIABLE section of the *Mplus* syntax, to conduct analyses on the similarity of profile solutions estimated independently in separate groups, the multiple group option needs to be activated, and defined using the KNOWNCLASS function. This uses a label (here, we use CG) to define this new grouping variable (in some variable-centered analyses, the multiple group commands are activated using the GROUPING command, but it is different for mixture models). In our example, we use two subgroups referring to men and women.

However, more than two groups can be used for any relevant subsample (e.g., cultural groups).

As noted, the levels of this grouping variable are defined as: (a) including participants with a value of 1 (men) on the variable Gender, and (b) including participants with a value of 2 (women) on the variable Gender (any other cases missing data on Gender or with an alternative

code will be omitted). There are now two grouping variables, C, the latent profiles, estimated as part of the LPA procedure, which have k levels (here, we are still working with a 3-profile solution), and CG, which reflects the *observed* subgroups (Gender) with 2 levels. Participants are allowed to be cross-classified.

```
VARIABLE:
KNOWNCLASS = CG(Gender = 1 Gender = 2);
CLASSES = CG(2) C(3);
```

The %OVERALL% section is then used to indicate that the class sizes (i.e., proportions of cases most likely associated to each profile) are freely estimated in all observed CG groups using the ON function, indicating that profile membership is conditional on gender. In *Mplus*, ON defines regressions (which are read as “outcome regressed *on* predictor”). Once again, $k-1$ statements are required (i.e., two multinomial intercepts of C are needed for a three-profile model). Profile-specific statements are then defined using combination labels of the known classes, CG, and the estimated classes, C. The ordering of the combined labels must reflect the ordering given in the CLASSES command of the VARIABLE section. For instance, if CG was listed first, it must occupy the first component of the label defined within the % signs. Similar labels in parentheses identify parameters that are constrained to equality across the CG groups. Here, none of the labels are shared between groups, so that the means and variances are freely estimated in all combinations of profiles across gender groups. Lists of constraints (m1–m3) apply to the parameters in order of appearance (m1 applies to IND1_T1, m2 to IND2_T1, m3 to IND3_T1).

Importantly, the syntax presented here *already assumes configural similarity*. More precisely, this syntax provides an example of how researchers can integrate a profile solution based only on men’s responses *with* a profile solution based only on women’s responses after

confirming that both solutions include the same number of latent profiles. This syntax and framework sets the stage for further tests of similarity (i.e., structural, dispersion) by adjusting the parameter labels given in parentheses. To determine if configural similarity is supported, researchers must conduct the profile enumeration procedure separately in each independent group to ensure that the optimal profile solution (i.e., with same number of profiles) is the same across groups. Configural similarity is the benchmark comparison model for the structural and other more restricted models (see Morin et al., 2016 for further details).

```
%OVERALL%
```

```
C#1 ON CG#1;
```

```
C#2 ON CG#1;
```

```
%CG#1.C#1%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m1-m3);
```

```
IND1_T1 IND2_T1 IND3_T1 (v1-v3);
```

```
%CG#1.C#2%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m4-m6);
```

```
IND1_T1 IND2_T1 IND3_T1 (v4-v6);
```

```
%CG#1.C#3%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m7-m9);
```

```
IND1_T1 IND2_T1 IND3_T1 (v7-v9);
```

```
%CG#2.C#1%
```

```
[IND1_T1 IND2_T1 IND3_T1] (mm1-mm3);
```

```
IND1_T1 IND2_T1 IND3_T1 (vv1-vv3);
```

```
%CG#2.C#2%
```

```
[IND1_T1 IND2_T1 IND3_T1] (mm4-mm6);
```

```
IND1_T1 IND2_T1 IND3_T1 (vv4-vv6);
```

```
%CG#2.C#3%
```

```
[IND1_T1 IND2_T1 IND3_T1] (mm7-mm9);
```

```
IND1_T1 IND2_T1 IND3_T1 (vv7-vv9);
```

Structural Similarity

The only difference between this model and the configural similarity model is that respective means, within respective profiles, are constrained to equality across gender groups using identical labels in parentheses. Labels in bold are those that were modified from the configural similarity model.

To support structural similarity, researchers should compare the information criteria (AIC, CAIC, BIC, and ABIC) from the structural model to those furnished by the configural model. As these indices are presented in “lower is better” form, the model with the lowest estimates provides a better fit and should be retained. However, as these indices are not always in agreement, Morin et al. (2016) suggested that so long as at least two of the AIC, CAIC, BIC or ABIC demonstrate the same or decreased value in a more restricted model (i.e., structural similarity vs. configural similarity) then the more restricted model can be supported. While we endorse this recommendation, we note that fit across profile similarity models is still an area of ongoing research for which best practice recommendations are likely to change. Additionally, a traditional likelihood ratio test can be computed using $-2 \times$ (loglikelihood value from structural

similarity model – loglikelihood from configural similarity model), which is distributed as χ^2 , with degrees of freedom equal to the difference in the number of free parameters estimated between configural and structural similarity models (see Eid, Langeheine, & Diener, 2003; Geiser, Lehmann, & Eid, 2006; Kankaraš, Moors, & Vermunt, 2011). However, these tests suffer the same limitations as traditional chi-square difference tests (i.e., over sensitivity to sample size and minor model misspecifications, etc.; Marsh, Hau, & Grayson, 2005) and are thus not recommended.

```
%OVERALL%
```

```
C#1 ON CG#1;
```

```
C#2 ON CG#1;
```

```
%CG#1.C#1%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m1-m3);
```

```
IND1_T1 IND2_T1 IND3_T1 (v1-v3);
```

```
%CG#1.C#2%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m4-m6);
```

```
IND1_T1 IND2_T1 IND3_T1 (v4-v6);
```

```
%CG#1.C#3%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m7-m9);
```

```
IND1_T1 IND2_T1 IND3_T1 (v7-v9);
```

```
%CG#2.C#1%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m1-m3);
```

IND1_T1 IND2_T1 IND3_T1 (vv1-vv3);

%CG#2.C#2%

[IND1_T1 IND2_T1 IND3_T1] (**m4-m6**);

IND1_T1 IND2_T1 IND3_T1 (vv4-vv6);

%CG#2.C#3%

[IND1_T1 IND2_T1 IND3_T1] (**m7-m9**);

IND1_T1 IND2_T1 IND3_T1 (vv7-vv9);

Dispersion Similarity

The only difference between this model and the structural similarity model is that respective variances are now constrained to equality across gender groups within respective profiles using identical labels in parentheses. Labels in bold are those that were modified from the structural similarity model.

%OVERALL%

C#1 ON CG#1;

C#2 ON CG#1;

%CG#1.C#1%

[IND1_T1 IND2_T1 IND3_T1] (m1-m3);

IND1_T1 IND2_T1 IND3_T1 (v1-v3);

%CG#1.C#2%

[IND1_T1 IND2_T1 IND3_T1] (m4-m6);

```
IND1_T1 IND2_T1 IND3_T1 (v4-v6);
```

```
%CG#1.C#3%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m7-m9);
```

```
IND1_T1 IND2_T1 IND3_T1 (v7-v9);
```

```
%CG#2.C#1%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m1-m3);
```

```
IND1_T1 IND2_T1 IND3_T1 (v1-v3);
```

```
%CG#2.C#2%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m4-m6);
```

```
IND1_T1 IND2_T1 IND3_T1 (v4-v6);
```

```
%CG#2.C#3%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m7-m9);
```

```
IND1_T1 IND2_T1 IND3_T1 (v7-v9);
```

Distribution Similarity

The only difference between this model and the dispersion similarity model is that the commands directly beneath the %OVERALL% section have been deactivated (or could be omitted entirely) to reflect the fact that the sizes of the profiles are no longer conditional on gender (or CG group membership).

```
%OVERALL%
```

```
! C#1 ON CG#1;
```

! C#2 ON CG#1;

%CG#1.C#1%

[IND1_T1 IND2_T1 IND3_T1] (m1-m3);

IND1_T1 IND2_T1 IND3_T1 (v1-v3);

%CG#1.C#2%

[IND1_T1 IND2_T1 IND3_T1] (m4-m6);

IND1_T1 IND2_T1 IND3_T1 (v4-v6);

%CG#1.C#3%

[IND1_T1 IND2_T1 IND3_T1] (m7-m9);

IND1_T1 IND2_T1 IND3_T1 (v7-v9);

%CG#2.C#1%

[IND1_T1 IND2_T1 IND3_T1] (m1-m3);

IND1_T1 IND2_T1 IND3_T1 (v1-v3);

%CG#2.C#2%

[IND1_T1 IND2_T1 IND3_T1] (m4-m6);

IND1_T1 IND2_T1 IND3_T1 (v4-v6);

%CG#2.C#3%

[IND1_T1 IND2_T1 IND3_T1] (m7-m9);

IND1_T1 IND2_T1 IND3_T1 (v7-v9);

MULTIPLE-GROUP LATENT PROFILE ANALYSES WITH PREDICTORS

Relations Freely Estimated Across Subgroups

This model uses the SVALUES associated with the most similar model retained from the previous sequence (here we start from a model of distributional similarity), specified with * followed by the value of the estimated parameters. To this model, the effects of predictors on profile membership can be incorporated (C#1-C#2 ON Pred1 Pred2;). To allow these effects to be freely estimated across gender, they need to be constrained and fixed to 0 in the %OVERALL% section, and then freely estimated in both gender groups in a new section of the input that specifically refers to CG. See all sections in bold.

```
%OVERALL%
```

```
[CG#1*-0.00217]; [C#1*-0.93515]; [C#2*0.49113];
```

```
C#1-C#2 ON Pred1@0 Pred2@0;
```

```
%CG#1.C#1%
```

```
[IND1_T1*-1.59367] (m1);
```

```
[IND2_T1*-0.93155] (m2);
```

```
[IND3_T1*0.39664] (m3);
```

```
IND1_T1*0.53789 (v1);
```

```
IND2_T1*2.12542 (v2);
```

```
IND3_T1*1.59137 (v3);
```

```
%CG#1.C#2%
```

```
[IND1_T1*-0.18248] (m4);
```

```
[IND2_T1*-0.29812] (m5);
```

[IND3_T1*-0.15201] (m6);

IND1_T1*0.42376 (v4);

IND2_T1*0.43116 (v5);

IND3_T1*0.72513 (v6);

%CG#1.C#3%

[IND1_T1*0.94111] (m7);

[IND2_T1*0.86526] (m8);

[IND3_T1*0.08333] (m9);

IND1_T1*0.44053 (v7);

IND2_T1*0.48030 (v8);

IND3_T1*1.13489 (v9);

%CG#2.C#1%

[IND1_T1*-1.59367] (m1);

[IND2_T1*-0.93155] (m2);

[IND3_T1*0.39664] (m3);

IND1_T1*0.53789 (v1);

IND2_T1*2.12542 (v2);

IND3_T1*1.59137 (v3);

%CG#2.C#2%

[IND1_T1*-0.18248] (m4);

IND2_T1*-0.29812] (m5);

[IND3_T1*-0.15201] (m6);

IND1_T1*0.42376 (v4);

IND2_T1*0.43116 (v5);

IND3_T1*0.72513 (v6);

%CG#2.C#3%

[IND1_T1*0.94111] (m7);

[IND2_T1*0.86526] (m8);

[IND3_T1*0.08333] (m9);

IND1_T1*0.44053 (v7);

IND2_T1*0.48030 (v8);

IND3_T1*1.13489 (v9);

MODEL CG:

%CG#1%

C#1-C#2 ON Pred1 Pred2;

%CG#2%

C#1-C#2 ON Pred1 Pred2;

Predictive Similarity

This model is almost identical to the previous freely estimated predictor effects model. In order for the predictor effects to be constrained to similarity across gender, they need to be specified as freely estimated in the %OVERALL% section (C#1-C#2 ON Pred1 Pred2;), while removing the gender-specific section (i.e., the MODEL CG: section). See the code in bold. Note also the carryover of using the SVALUES as associated with the distributional similarity model.

%OVERALL%

[CG#1*-0.00217]; [C#1*-0.93515]; [C#2*0.49113];

C#1-C#2 ON Pred1 Pred2;

%CG#1.C#1%

[IND1_T1*-1.59367] (m1);

[IND2_T1*-0.93155] (m2);

[IND3_T1*0.39664] (m3);

IND1_T1*0.53789 (v1);

IND2_T1*2.12542 (v2);

IND3_T1*1.59137 (v3);

%CG#1.C#2%

[IND1_T1*-0.18248] (m4);

[IND2_T1*-0.29812] (m5);

[IND3_T1*-0.15201] (m6);

IND1_T1*0.42376 (v4);

IND2_T1*0.43116 (v5);

IND3_T1*0.72513 (v6);

%CG#1.C#3%

[IND1_T1*0.94111] (m7);

[IND2_T1*0.86526] (m8);

[IND3_T1*0.08333] (m9);

IND1_T1*0.44053 (v7);

IND2_T1*0.48030 (v8);

IND3_T1*1.13489 (v9);

%CG#2.C#1%

[IND1_T1*-1.59367] (m1);

[IND2_T1*-0.93155] (m2);

[IND3_T1*0.39664] (m3);

IND1_T1*0.53789 (v1);

IND2_T1*2.12542 (v2);

IND3_T1*1.59137 (v3);

%CG#2.C#2%

[IND1_T1*-0.18248] (m4);

IND2_T1*-0.29812] (m5);

[IND3_T1*-0.15201] (m6);

IND1_T1*0.42376 (v4);

IND2_T1*0.43116 (v5);

IND3_T1*0.72513 (v6);

%CG#2.C#3%

[IND1_T1*0.94111] (m7);

[IND2_T1*0.86526] (m8);

[IND3_T1*0.08333] (m9);

```

IND1_T1*0.44053 (v7);
IND2_T1*0.48030 (v8);
IND3_T1*1.13489 (v9);

```

MULTIPLE-GROUP LATENT PROFILE ANALYSES WITH OUTCOMES

Relations Freely Estimated Across Subgroups

This model also uses the SVALUES associated with the distributional similarity model. Similar to the direct inclusion and manual implementation approaches to covariates, we request the free estimation of the distal outcome means and variances in all profiles across gender (i.e., [Out1]; Out1; [Out2]; Out2;), and add labels in parentheses to identify these new parameters. These new parameters will then be used in a MODEL CONSTRAINT section to request tests of the significance of mean differences between profiles within genders. See the code in bold.

```

%OVERALL%

[CG#1*-0.00217]; [C#1*-0.93515]; [C#2*0.49113];

%CG#1.C#1%

[IND1_T1*-1.59367] (m1);
[IND2_T1*-0.93155] (m2);
[IND3_T1*0.39664] (m3);
IND1_T1*0.53789 (v1);
IND2_T1*2.12542 (v2);
IND3_T1*1.59137 (v3);

[Out1] (oa1);

Out1;

[Out2] (ob1);

```

Out2;

%CG#1.C#2%

[IND1_T1*-0.18248] (m4);

[IND2_T1*-0.29812] (m5);

[IND3_T1*-0.15201] (m6);

IND1_T1*0.42376 (v4);

IND2_T1*0.43116 (v5);

IND3_T1*0.72513 (v6);

[Out1] (oa2);**Out1;****[Out2] (ob2);****Out2;**

%CG#1.C#3%

[IND1_T1*0.94111] (m7);

[IND2_T1*0.86526] (m8);

[IND3_T1*0.08333] (m9);

IND1_T1*0.44053 (v7);

IND2_T1*0.48030 (v8);

IND3_T1*1.13489 (v9);

[Out1] (oa3);**Out1;****[Out2] (ob3);****Out2;**

%CG#2.C#1%

[IND1_T1*-1.59367] (m1);

[IND2_T1*-0.93155] (m2);

[IND3_T1*0.39664] (m3);

IND1_T1*0.53789 (v1);

IND2_T1*2.12542 (v2);

IND3_T1*1.59137 (v3);

[Out1] (oaa1);

Out1;

[Out2] (obb1);

Out2;

%CG#2.C#2%

[IND1_T1*-0.18248] (m4);

IND2_T1*-0.29812] (m5);

[IND3_T1*-0.15201] (m6);

IND1_T1*0.42376 (v4);

IND2_T1*0.43116 (v5);

IND3_T1*0.72513 (v6);

[Out1] (oaa2);

Out1;

[Out2] (obb2);

Out2;

%CG#2.C#3%

[IND1_T1*0.94111] (m7);

[IND2_T1*0.86526] (m8);

[IND3_T1*0.08333] (m9);

IND1_T1*0.44053 (v7);

```

IND2_T1*0.48030 (v8);
IND3_T1*1.13489 (v9);

[Out1] (oaa3);
Out1;

[Out2] (obb3);
Out2;

MODEL CONSTRAINT:

NEW (y12); y12 = oa1-oa2;
NEW (y13); y13 = oa1-oa3;
NEW (y23); y23 = oa2-oa3;
NEW (z12); z12 = ob1-ob2
NEW (z13); z13 = ob1-ob3;
NEW (z23); z23 = ob2-ob3;

NEW (yy12); yy12 = oaa1-oaa2;
NEW (yy13); yy13 = oaa1-oaa3;
NEW (yy23); yy23 = oaa2-oaa3;
NEW (zz12); zz12 = obb1-obb2;
NEW (zz13); zz13 = obb1-obb3;
NEW (zz23); zz23 = obb2-obb3;

```

Explanatory Similarity

This model is almost identical to the previous model, except that the parameter labels are used to constrain the outcome means are invariant across gender. As a result, fewer lines of code are required in the MODEL CONSTRAINT section.

%OVERALL%

[CG#1*-0.00217]; [C#1*-0.93515]; [C#2*0.49113];

%CG#1.C#1%

[IND1_T1*-1.59367] (m1);

[IND2_T1*-0.93155] (m2);

[IND3_T1*0.39664] (m3);

IND1_T1*0.53789 (v1);

IND2_T1*2.12542 (v2);

IND3_T1*1.59137 (v3);

[Out1] (oa1);

Out1;

[Out2] (ob1);

Out2;

%CG#1.C#2%

[IND1_T1*-0.18248] (m4);

[IND2_T1*-0.29812] (m5);

[IND3_T1*-0.15201] (m6);

IND1_T1*0.42376 (v4);

IND2_T1*0.43116 (v5);

IND3_T1*0.72513 (v6);

[Out1] (oa2);

Out1;

[Out2] (ob2);

Out2;

%CG#1.C#3%

[IND1_T1*0.94111] (m7);

[IND2_T1*0.86526] (m8);

[IND3_T1*0.08333] (m9);

IND1_T1*0.44053 (v7);

IND2_T1*0.48030 (v8);

IND3_T1*1.13489 (v9);

[Out1] (oa3);

Out1;

[Out2] (ob3);

Out2;

%CG#2.C#1%

[IND1_T1*-1.59367] (m1);

[IND2_T1*-0.93155] (m2);

[IND3_T1*0.39664] (m3);

IND1_T1*0.53789 (v1);

IND2_T1*2.12542 (v2);

IND3_T1*1.59137 (v3);

[Out1] (oa1);

Out1;

[Out2] (ob1);

Out2;

%CG#2.C#2%

[IND1_T1*-0.18248] (m4);

IND2_T1*-0.29812] (m5);

[IND3_T1*-0.15201] (m6);

IND1_T1*0.42376 (v4);

IND2_T1*0.43116 (v5);

IND3_T1*0.72513 (v6);

[Out1] (oa2);

Out1;

[Out2] (ob2);

Out2;

%CG#2.C#3%

[IND1_T1*0.94111] (m7);

[IND2_T1*0.86526] (m8);

[IND3_T1*0.08333] (m9);

IND1_T1*0.44053 (v7);

IND2_T1*0.48030 (v8);

IND3_T1*1.13489 (v9);

[Out1] (oa3);

Out1;

[Out2] (ob3);

Out2;

MODEL CONSTRAINT:

NEW (y12); y12 = oa1-oa2;

NEW (y13); y13 = oa1-oa3;

NEW (y23); y23 = oa2-oa3;

NEW (z12); z12 = ob1-ob2;

NEW (z13); z13 = ob1-ob3;

NEW (z23); z23 = ob2-ob3;

MIXTURE REGRESSION ANALYSIS

Mixture regression analyses specify a regression model in the %OVERALL% section of the input indicating that an outcome, Out1, is regressed ON a series of predictors, IND1_T1, IND2_T1, and IND3_T1. Then, the profile-specific sections of the input request that these regression relations be freely estimated across each profile. In the basic mixture regression model, the mean and variance of the outcome(s) also need to be freely estimated in each profile as they respectively reflect the intercepts and residuals of the regression equations. One aspect of this basic model is that the means and variances of the predictors are held equal across classes because they are considered exogenous to the focal mixture (which reflects the underlying regression equation).

%OVERALL%

Out1 ON IND1_T1 IND2_T1 IND3_T1;

%C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1;

Out1; [Out1];

```

% C#2 %
Out1 ON IND1_T1 IND2_T1 IND3_T1;

Out1; [Out1];

% C#3 %
Out1 ON IND1_T1 IND2_T1 IND3_T1;

Out1; [Out1];

```

A more flexible (and potentially more realistic) representation of a mixture regression model also specifies freely estimated means and variances of the predictors in each profile. This results in a hybrid model that combines LPA (for the predictors) and mixture regression(s), and can highlight how the regression relations differ as a function of the latent profiles that are defined on the basis of differential configurations of predictor scores (Chénard Poirier, Morin, & Boudrias, 2017; see also McLarnon & O’Neill, 2018). As above, this specification results in correlations between the exogenous variables (i.e., IND1_T1, IND2_T1, and IND3_T1) that are constrained to equality across profiles. These constraints, and their implied homogeneity assumption, can be relaxed by additionally including IND1_T1 WITH IND2_T1 IND3_T1; IND2_T1 WITH IND3_T1; in each profile-specific subsection. Though, given concerns highlighted previously, caution should be exercised in implementing these changes and incorporating correlated uniquenesses.

```

% OVERALL %
Out1 ON IND1_T1 IND2_T1 IND3_T1;

% C#1 %
Out1 ON IND1_T1 IND2_T1 IND3_T1;

```

```
[Out1 IND1_T1 IND2_T1 IND3_T1];
Out1 IND1_T1 IND2_T1 IND3_T1;

%C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1;
[Out1 IND1_T1 IND2_T1 IND3_T1];
Out1 IND1_T1 IND2_T1 IND3_T1;

%C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1;
[Out1 IND1_T1 IND2_T1 IND3_T1];
Out1 IND1_T1 IND2_T1 IND3_T1;
```

MULTIPLE GROUP MIXTURE REGRESSION ANALYSIS

Configural Similarity

This set-up is highly similar to the one used for the multiple groups LPA models. Here again, the `KNOWNCLASS` option is used to define the gender groups in the `VARIABLE` section:

```
KNOWNCLASS = CG(Gender = 1 Gender = 2);
CLASSES = CG(2) C(3);
```

The `MODEL` then section describes the model of configural similarity.

```
%OVERALL%
C#1 ON CG#1;
C#2 ON CG#1;
Out1 ON IND1_T1 IND2_T1 IND3_T1;
```

%CG#1.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);

Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);

%CG#1.C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);

Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);

%CG#1.C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);

Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);

%CG#2.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (rr1-rr3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (mm1-mm4);

Out1 IND1_T1 IND2_T1 IND3_T1 (vv1-vv4);

%CG#2.C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (rr11-rr13);

[Out1 IND1_T1 IND2_T1 IND3_T1] (mm11-mm14);

Out1 IND1_T1 IND2_T1 IND3_T1 (vv11-vv14);

```
%CG#2.C#3%
```

```
Out1 ON IND1_T1 IND2_T1 IND3_T1 (rr21-rr23);
```

```
[Out1 IND1_T1 IND2_T1 IND3_T1] (mm21-mm24);
```

```
Out1 IND1_T1 IND2_T1 IND3_T1 (vv21-vv24);
```

Regression Similarity

The only difference between this model and the previous one is that the regression coefficients are now constrained to equality across gender groups within respective profiles using identical parameter labels in parentheses. Labels in bold were modified from the previous model.

```
%OVERALL%
```

```
C#1 ON CG#1;
```

```
C#2 ON CG#1;
```

```
Out1 ON IND1_T1 IND2_T1 IND3_T1;
```

```
%CG#1.C#1%
```

```
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);
```

```
[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);
```

```
Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);
```

```
%CG#1.C#2%
```

```
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);
```

```
[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);
```

```
Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);
```

%CG#1.C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (**r21-r23**);

[Out1 IND1_T1 IND2_T1 IND3_T1] (**m21-m24**);

Out1 IND1_T1 IND2_T1 IND3_T1 (**v21-v24**);

%CG#2.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (**r1-r3**);

[Out1 IND1_T1 IND2_T1 IND3_T1] (**mm1-mm4**);

Out1 IND1_T1 IND2_T1 IND3_T1 (**vv1-vv4**);

%CG#2.C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (**r11-r13**);

[Out1 IND1_T1 IND2_T1 IND3_T1] (**mm11-mm14**);

Out1 IND1_T1 IND2_T1 IND3_T1 (**vv11-vv14**);

%CG#2.C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (**r21-r23**);

[Out1 IND1_T1 IND2_T1 IND3_T1] (**mm21-mm24**);

Out1 IND1_T1 IND2_T1 IND3_T1 (**vv21-vv24**);

Structural Similarity

Means of both predictors and outcome are constrained to equality across gender within each mixture group using identical labels in parentheses. Labels in bold were modified from the previous model.

%OVERALL%

C#1 ON CG#1;

C#2 ON CG#1;

Out1 ON IND1_T1 IND2_T1 IND3_T1;

%CG#1.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);

Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);

%CG#1.C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);

Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);

%CG#1.C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);

Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);

%CG#2.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (**m1-m4**);

Out1 IND1_T1 IND2_T1 IND3_T1 (vv1-vv4);

```

%CG#2.C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);

Out1 IND1_T1 IND2_T1 IND3_T1 (vv11-vv14);

%CG#2.C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);

Out1 IND1_T1 IND2_T1 IND3_T1 (vv21-vv24);

```

Dispersion Similarity

Variances are constrained to equality across gender within each profile using identical labels in parentheses. Labels in bold were modified from the previous model.

```

%OVERALL%

C#1 ON CG#1;

C#2 ON CG#1;

Out1 ON IND1_T1 IND2_T1 IND3_T1;

%CG#1.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);

Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);

%CG#1.C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);

```

[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);

Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);

%CG#1.C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);

Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);

%CG#2.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);

Out1 IND1_T1 IND2_T1 IND3_T1 (**v1-v4**);

%CG#2.C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);

Out1 IND1_T1 IND2_T1 IND3_T1 (**v11-v14**);

%CG#2.C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);

Out1 IND1_T1 IND2_T1 IND3_T1 (**v21-v24**);

Distributional Similarity

The only difference between this model and the dispersion similarity mixture regression model

is that the commands associated with the $C\#k$ ON $CG\#k$ regressions within the %OVERALL% section have been omitted to reflect that the sizes of the mixture groups are no longer conditional on gender (or group membership).

%OVERALL%

Out1 ON IND1_T1 IND2_T1 IND3_T1;

%CG#1.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);

Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);

%CG#1.C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);

Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);

%CG#1.C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);

Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);

%CG#2.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);

Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);

```
%CG#2.C#2%
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);
[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);
Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);
```

```
%CG#2.C#3%
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);
[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);
Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);
```

MULTIPLE GROUP MIXTURE REGRESSION ANALYSIS WITH PREDICTORS

Relations Freely Estimated Across Subgroups

To ensure stability, starting values from the previously most similar solution should be used (not illustrated here for clarity). To allow the predictor relations to be freely estimated across gender, they need to be constrained to 0 in the %OVERALL% section, and freely estimated in both gender groups in a new section of the input specifically referring to MODEL CG: (see commands in bold).

```
%OVERALL%
Out1 ON IND1_T1 IND2_T1 IND3_T1;
C#1-C#2 ON Pred1@0 Pred2@0;

%CG#1.C#1%
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);
[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);
```

Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);
%CG#1.C#2%
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);
[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);
Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);
%CG#1.C#3%
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);
[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);
Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);

%CG#2.C#1%
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);
[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);
Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);
%CG#2.C#2%
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);
[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);
Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);
%CG#2.C#3%
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);
[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);
Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);

MODEL CG:

%CG#1%

```

C#1-C#2 ON Pred1 Pred2;

%CG#2%

C#1-C#2 ON Pred1 Pred2;

```

Predictive Similarity

In order for the effects of the predictors to be constrained to similarity across gender, they simply need to be specified as freely estimated in the %OVERALL% section (C#1-C#2 ON Pred1 Pred2;), in conjunction with removing the gender specific section (i.e., the MODEL CG: section). See the code in bold. Again, to ensure stability, starting values from the previously most similar solution should be used.

```

%OVERALL%

Out1 ON IND1_T1 IND2_T1 IND3_T1;

C#1-C#2 ON Pred1 Pred2;

%CG#1.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);
[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);
Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);

%CG#1.C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);
[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);
Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);

%CG#1.C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);

```

```
[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);
```

```
Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);
```

```
%CG#2.C#1%
```

```
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);
```

```
[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);
```

```
Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);
```

```
%CG#2.C#2%
```

```
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);
```

```
[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);
```

```
Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);
```

```
%CG#2.C#3%
```

```
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);
```

```
[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);
```

```
Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);
```

MULTIPLE GROUP MIXTURE REGRESSION ANALYSIS WITH OUTCOMES

Relations Freely Estimated Across Subgroups

Here, we request the free estimation of the distal outcome means ([Out2];) in all profile groups across gender. We also use labels in parentheses to identify these parameters, which are then used in a MODEL CONSTRAINT section to request tests of the significance of the mean differences. See the code in bold. To ensure stability, start values from the previously most similar solution should be used.

%OVERALL%

Out1 ON IND1_T1 IND2_T1 IND3_T1;

%CG#1.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);

Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);

[Out2] (ob1); Out2;

%CG#1.C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);

Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);

[Out2] (ob2); Out2;

%CG#1.C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);

Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);

[Out2] (ob3); Out2;

%CG#2.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);

Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);

[Out2] (obb1); Out2;

%CG#2.C#2%

```
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);
```

```
[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);
```

```
Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);
```

```
[Out2] (obb2); Out2;
```

```
%CG#2.C#3%
```

```
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);
```

```
[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);
```

```
Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);
```

```
[Out2] (obb3); Out2;
```

```
MODEL CONSTRAINT:
```

```
NEW (z12); z12 = ob1-ob2;
```

```
NEW (z13); z13 = ob1-ob3;
```

```
NEW (z23); z23 = ob2-ob3;
```

```
NEW (zz12); zz12 = obb1-obb2;
```

```
NEW (zz13); zz13 = obb1-obb3;
```

```
NEW (zz23); zz23 = obb2-obb3;
```

Explanatory Similarity

This model is almost identical to the previous model, except that parameter labels are used to constrain the outcome means to be invariant across gender. As a result, fewer lines of code are required in the MODEL COSNTRAIINT section. To ensure stability, starting values from the previously most similar solution should be used.

%OVERALL%

Out1 ON IND1_T1 IND2_T1 IND3_T1;

%CG#1.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);

Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);

[Out2] (ob1); Out2;

%CG#1.C#2%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);

Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);

[Out2] (ob2); Out2;

%CG#1.C#3%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);

Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);

[Out2] (ob3); Out2;

%CG#2.C#1%

Out1 ON IND1_T1 IND2_T1 IND3_T1 (r1-r3);

[Out1 IND1_T1 IND2_T1 IND3_T1] (m1-m4);

Out1 IND1_T1 IND2_T1 IND3_T1 (v1-v4);

[Out2] (ob1); Out2;

%CG#2.C#2%

```
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r11-r13);
```

```
[Out1 IND1_T1 IND2_T1 IND3_T1] (m11-m14);
```

```
Out1 IND1_T1 IND2_T1 IND3_T1 (v11-v14);
```

```
[Out2] (ob2); Out2;
```

```
%CG#2.C#3%
```

```
Out1 ON IND1_T1 IND2_T1 IND3_T1 (r21-r23);
```

```
[Out1 IND1_T1 IND2_T1 IND3_T1] (m21-m24);
```

```
Out1 IND1_T1 IND2_T1 IND3_T1 (v21-v24);
```

```
[Out2] (ob3); Out2;
```

```
MODEL CONSTRAINT:
```

```
NEW (z12); z12 = ob1-ob2;
```

```
NEW (z13); z13 = ob1-ob3;
```

```
NEW (z23); z23 = ob2-ob3;
```

LONGITUDINAL LATENT PROFILE ANALYSIS

Configural Similarity

The estimation of a longitudinal latent profile analysis is highly similar to the estimation of a multiple-group LPA with the exception that the KNOWNCLASS function (which previously defined the observed grouping variable, gender, and was coded as CG) is replaced with another unobserved latent categorical variable that represents profiles groups estimated at another time point. The VARIABLE section of the syntax is of course also adapted to include both sets of Time 1 and Time 2 indicators).

```
CLASSES = C1(3) C2(3);
```

Because of the similarity of the inputs, we do not comment on the sequence of similarity tests in the next sections (see commands in bold that denote components of the syntax that are adjusted at each step), and refer readers to the multi-group models. In this model, two sections of the input are used to define the profiles estimated at the first (MODEL C1:) and second (MODEL C2:) time points, where the profiles are defined by distinct variables reflecting the indicators measured at either the first (e.g., IND1_T1) or second (e.g., IND1_T2) time point.

```
%OVERALL%
```

```
MODEL C1:
```

```
%C1#1%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m1-m3);
```

```
IND1_T1 IND2_T1 IND3_T1 (v1-v3);
```

```
%C1#2%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m4-m6);
```

```
IND1_T1 IND2_T1 IND3_T1 (v4-v6);
```

```
%C1#3%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m7-m9);
```

```
IND1_T1 IND2_T1 IND3_T1 (v7-v9);
```

```
MODEL C2:
```

```
%C2#1%
```

```
[IND1_T2 IND2_T2 IND3_T2] (mm1-mm3);
```

```
IND1_T2 IND2_T2 IND3_T2 (vv1-vv3);
```

```
%C2#2%
```

```
[IND1_T2 IND2_T2 IND3_T2] (mm4-mm6);
```

```
IND1_T2 IND2_T2 IND3_T2 (vv4-vv6);
```

```
%C2#3%
```

[IND1_T2 IND2_T2 IND3_T2] (mm7-mm9);

IND1_T2 IND2_T2 IND3_T2 (vv7-vv9);

Structural Similarity

%OVERALL%

MODEL C1:

%C1#1%

[IND1_T1 IND2_T1 IND3_T1] (m1-m3);

IND1_T1 IND2_T1 IND3_T1 (v1-v3);

%C1#2%

[IND1_T1 IND2_T1 IND3_T1] (m4-m6);

IND1_T1 IND2_T1 IND3_T1 (v4-v6);

%C1#3%

[IND1_T1 IND2_T1 IND3_T1] (m7-m9);

IND1_T1 IND2_T1 IND3_T1 (v7-v9);

MODEL C2:

%C2#1%

[IND1_T2 IND2_T2 IND3_T2] (**m1-m3**);

IND1_T2 IND2_T2 IND3_T2 (vv1-vv3);

%C2#2%

[IND1_T2 IND2_T2 IND3_T2] (**m4-m6**);

IND1_T2 IND2_T2 IND3_T2 (vv4-vv6);

%C2#3%

[IND1_T2 IND2_T2 IND3_T2] (**m7-m9**);

IND1_T2 IND2_T2 IND3_T2 (vv7-vv9);

Dispersion Similarity

%OVERALL%

MODEL C1:

%C1#1%

[IND1_T1 IND2_T1 IND3_T1] (m1-m3);

IND1_T1 IND2_T1 IND3_T1 (v1-v3);

%C1#2%

[IND1_T1 IND2_T1 IND3_T1] (m4-m6);

IND1_T1 IND2_T1 IND3_T1 (v4-v6);

%C1#3%

[IND1_T1 IND2_T1 IND3_T1] (m7-m9);

IND1_T1 IND2_T1 IND3_T1 (v7-v9);

MODEL C2:

%C2#1%

[IND1_T2 IND2_T2 IND3_T2] (m1-m3);

IND1_T2 IND2_T2 IND3_T2 (v1-v3);

%C2#2%

[IND1_T2 IND2_T2 IND3_T2] (m4-m6);

IND1_T2 IND2_T2 IND3_T2 (v4-v6);

%C2#3%

[IND1_T2 IND2_T2 IND3_T2] (m7-m9);

IND1_T2 IND2_T2 IND3_T2 (v7-v9);

Distributional Similarity

In contrast to the multi-group distributional similarity model, which invoked regressions of the

$C\#k$ profiles on the $CG\#k$ groupings, in longitudinal LPAs, labels are used to constrain the sizes of the profiles (proportions of cases assigned to each profile) to equality over time points.

%OVERALL%

[C1#1] (p1);

[C1#2] (p2);

[C2#1] (p1);

[C2#2] (p2);

MODEL C1:

%C1#1%

[IND1_T1 IND2_T1 IND3_T1] (m1-m3);

IND1_T1 IND2_T1 IND3_T1 (v1-v3);

%C1#2%

[IND1_T1 IND2_T1 IND3_T1] (m4-m6);

IND1_T1 IND2_T1 IND3_T1 (v4-v6);

%C1#3%

[IND1_T1 IND2_T1 IND3_T1] (m7-m9);

IND1_T1 IND2_T1 IND3_T1 (v7-v9);

MODEL C2:

%C2#1%

[IND1_T2 IND2_T2 IND3_T2] (m1-m3);

IND1_T2 IND2_T2 IND3_T2 (v1-v3);

```
%C2#2%
```

```
[IND1_T2 IND2_T2 IND3_T2] (m4-m6);
```

```
IND1_T2 IND2_T2 IND3_T2 (v4-v6);
```

```
%C2#3%
```

```
[IND1_T2 IND2_T2 IND3_T2] (m7-m9);
```

```
IND1_T2 IND2_T2 IND3_T2 (v7-v9);
```

LATENT TRANSITION ANALYSIS (FROM AN LPA MODEL OF DISPERSION SIMILARITY)

In the basic two-occasion LTA, the %OVERALL% section states that membership into the profiles at the second time point (C2) are conditional on membership in the profiles estimated at the first time point (C1). This is necessary to estimate the individual transition probabilities over time. This is the only change (see command in bold) that is required to convert a longitudinal LPA model with dispersional similarity to an LTA model. However, an important caveat is that the procedure required to convert an LPA model with distributional similarity to a LTA is more complex (for details, see Morin & Litalien, 2017: <https://smslabstats.weebly.com/webnotes.html>).

```
%OVERALL%
```

```
C2 ON C1;
```

```
MODEL C1:
```

```
%C1#1%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m1-m3);
```

IND1_T1 IND2_T1 IND3_T1 (v1-v3);

%C1#2%

[IND1_T1 IND2_T1 IND3_T1] (m4-m6);

IND1_T1 IND2_T1 IND3_T1 (v4-v6);

%C1#3%

[IND1_T1 IND2_T1 IND3_T1] (m7-m9);

IND1_T1 IND2_T1 IND3_T1 (v7-v9);

MODEL C2:

%C2#1%

[IND1_T2 IND2_T2 IND3_T2] (m1-m3);

IND1_T2 IND2_T2 IND3_T2 (v1-v3);

%C2#2%

[IND1_T2 IND2_T2 IND3_T2] (m4-m6);

IND1_T2 IND2_T2 IND3_T2 (v4-v6);

%C2#3%

[IND1_T2 IND2_T2 IND3_T2] (m7-m9);

IND1_T2 IND2_T2 IND3_T2 (v7-v9);

LATENT TRANSITION ANALYSIS WITH PREDICTORS

Relations Freely Estimated Across Profiles and Time Points

In this model, the effects of the predictors are allowed to differ across time points. Specifically,

separate commands are given for the regression of Time 1 and Time 2 profiles (C1 and C2, respectively) on the predictors. In addition, the effects of the predictors on the Time 2 profiles are allowed to differ across the Time 1 profiles in order to assess their specific role in the prediction of specific profile-to-profile transitions. In other words, this allows for the predictors to moderate the transition probabilities between C1 and C2 profiles. As in previous models, to ensure stability, starting values from the previously most similar solution should be used. In conjunction with this parameterization, users may wish to add TECH15 to the OUTPUT command to obtain marginal and conditional probabilities (including transition probabilities) for the latent categorical (i.e., profiles) variables included in the model. As well, conditional probabilities of individual profile membership and transitions probabilities can be computed for different values (i.e., mean, $\pm 1SD$; see McLarnon et al., 2019) of the predictors to assess presence of a moderating effect of the predictors using *Mplus*' built-in LTA calculator (which is available from the *Mplus* drop-down menu of the *Mplus* Editor window).

```
%OVERALL%
```

```
C2 on C1;
```

```
[C1#1] (p1); [C1#2] (p2);
```

```
[C2#1] (p1); [C2#2] (p2);
```

```
C1 ON Pred1 Pred2;
```

```
C2 ON Pred1 Pred2;
```

```
MODEL C1:
```

```
%C1#1%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m1-m3);
```

```
IND1_T1 IND2_T1 IND3_T1 (v1-v3);
```

```
C2 ON Pred1 Pred2;
```

%C1#2%

[IND1_T1 IND2_T1 IND3_T1] (m4-m6);

IND1_T1 IND2_T1 IND3_T1 (v4-v6);

C2 ON Pred1 Pred2;

%C1#3%

[IND1_T1 IND2_T1 IND3_T1] (m7-m9);

IND1_T1 IND2_T1 IND3_T1 (v7-v9);

C2 ON Pred1 Pred2;

MODEL C2:

%C2#1%

[IND1_T2 IND2_T2 IND3_T2] (m1-m3);

IND1_T2 IND2_T2 IND3_T2 (v1-v3);

%C2#2%

[IND1_T2 IND2_T2 IND3_T2] (m4-m6);

IND1_T2 IND2_T2 IND3_T2 (v4-v6);

%C2#3%

[IND1_T2 IND2_T2 IND3_T2] (m7-m9);

IND1_T2 IND2_T2 IND3_T2 (v7-v9);

OUTPUT: TECH15;

Relations Freely Estimated Across Time Points

This model is very similar to the previous one, and has only removed the commands that specified unique predictor relations across the Time 1 profiles. To ensure stability, start values from the previously most similar solution should be used.

```
%OVERALL%
```

```
C2 on C1;
```

```
[C1#1] (p1);
```

```
[C1#2] (p2);
```

```
[C2#1] (p1);
```

```
[C2#2] (p2);
```

```
C1 ON Pred1 Pred2;
```

```
C2 ON Pred1 Pred2;
```

```
MODEL C1:
```

```
%C1#1%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m1-m3);
```

```
IND1_T1 IND2_T1 IND3_T1 (v1-v3);
```

```
%C1#2%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m4-m6);
```

```
IND1_T1 IND2_T1 IND3_T1 (v4-v6);
```

```
%C1#3%
```

```
[IND1_T1 IND2_T1 IND3_T1] (m7-m9);
```

IND1_T1 IND2_T1 IND3_T1 (v7-v9);

MODEL C2:

%C2#1%

[IND1_T2 IND2_T2 IND3_T2] (m1-m3);

IND1_T2 IND2_T2 IND3_T2 (v1-v3);

%C2#2%

[IND1_T2 IND2_T2 IND3_T2] (m4-m6);

IND1_T2 IND2_T2 IND3_T2 (v4-v6);

%C2#3%

[IND1_T2 IND2_T2 IND3_T2] (m7-m9);

IND1_T2 IND2_T2 IND3_T2 (v7-v9);

Predictive Similarity

To ensure stability, start values from the previously most similar solution should be used.

%OVERALL%

C2 on C1;

[C1#1] (p1);

[C1#2] (p2);

[C2#1] (p1);

[C2#2] (p2);

C1 ON Pred1 (pr1-pr2);

! one fewer labels than the number of latent profiles is needed

! i.e., the full expansion of the parameters implied here is:

! C1#1 ON Pred1 (pr1); C1#2 ON Pred1 (pr2);

C2 ON Pred1 (pr1-pr2);

C1 ON Pred2 (pr3-pr4);

C2 ON Pred2 (pr3-pr4);

MODEL C1:

%C1#1%

[IND1_T1 IND2_T1 IND3_T1] (m1-m3);

IND1_T1 IND2_T1 IND3_T1 (v1-v3);

%C1#2%

[IND1_T1 IND2_T1 IND3_T1] (m4-m6);

IND1_T1 IND2_T1 IND3_T1 (v4-v6);

%C1#3%

[IND1_T1 IND2_T1 IND3_T1] (m7-m9);

IND1_T1 IND2_T1 IND3_T1 (v7-v9);

MODEL C2:

%C2#1%

[IND1_T2 IND2_T2 IND3_T2] (m1-m3);

IND1_T2 IND2_T2 IND3_T2 (v1-v3);

```
%C2#2%
```

```
[IND1_T2 IND2_T2 IND3_T2] (m4-m6);
```

```
IND1_T2 IND2_T2 IND3_T2 (v4-v6);
```

```
%C2#3%
```

```
[IND1_T2 IND2_T2 IND3_T2] (m7-m9);
```

```
IND1_T2 IND2_T2 IND3_T2 (v7-v9);
```

LATENT TRANSITION ANALYSIS WITH OUTCOMES

Relations Freely Estimated Across Time Points

Here, we use a case where a single outcome, Out, was measured twice, and show how time-invariant relations could be estimated (i.e., Out1 is the outcome for only the Time 1 profiles, and Out2 is the same outcome but measured at Time 2). As a separate research question, mean differences across the different transition patterns could be investigated by applying similarly structured syntax as described for the multi-group LPA (i.e., develop %C1#1.C2#1% and provide [Out1]; Out1; commands to allow the outcome to vary across transition patterns. As well, parameter labels can be applied to each of these outcome means, and then tested for significant differences using a series of MODEL CONSTRAINTS as illustrated here; see McLarnon et al., 2019 for an example and syntax). To ensure stability, start values from the previously most similar solution should be used.

```
%OVERALL%
```

```
C2 on C1;
```

```
[C1#1] (p1);
```

```
[C1#2] (p2);
```

[C2#1] (p1);

[C2#2] (p2);

MODEL C1:

%C1#1%

[IND1_T1 IND2_T1 IND3_T1] (m1-m3);

IND1_T1 IND2_T1 IND3_T1 (v1-v3);

[Out1] (oa1); Out1;

%C1#2%

[IND1_T1 IND2_T1 IND3_T1] (m4-m6);

IND1_T1 IND2_T1 IND3_T1 (v4-v6);

[Out1] (oa2); Out1;

%C1#3%

[IND1_T1 IND2_T1 IND3_T1] (m7-m9);

IND1_T1 IND2_T1 IND3_T1 (v7-v9);

[Out1] (oa3); Out1;

MODEL C2:

%C2#1%

[IND1_T2 IND2_T2 IND3_T2] (m1-m3);

IND1_T2 IND2_T2 IND3_T2 (v1-v3);

[Out2] (ob1); Out2;

%C2#2%

[IND1_T2 IND2_T2 IND3_T2] (m4-m6);

IND1_T2 IND2_T2 IND3_T2 (v4-v6);

[Out2] (ob2); Out2;

%C2#3%

[IND1_T2 IND2_T2 IND3_T2] (m7-m9);

IND1_T2 IND2_T2 IND3_T2 (v7-v9);

[Out2] (ob3); Out2;

MODEL CONSTRAINT:

NEW (y12); y12 = oa1-oa2;

NEW (y13); y13 = oa1-oa3;

NEW (y23); y23 = oa2-oa3;

NEW (z12); z12 = ob1-ob2

NEW (z13); z13 = ob1-ob3;

NEW (z23); z23 = ob2-ob3;

Explanatory Similarity

Here, similar labels for the outcomes are applied across Time 1 and Time 2 profiles. As noted previously, to ensure stability, starting values from the previously most similar solution should be used.

%OVERALL%

C2 on C1;

[C1#1] (p1);

[C1#2] (p2);

[C2#1] (p1);

[C2#2] (p2);

MODEL C1:

%C1#1%

[IND1_T1 IND2_T1 IND3_T1] (m1-m3);

IND1_T1 IND2_T1 IND3_T1 (v1-v3);

[Out1] (oa1); Out1;

%C1#2%

[IND1_T1 IND2_T1 IND3_T1] (m4-m6);

IND1_T1 IND2_T1 IND3_T1 (v4-v6);

[Out1] (oa2); Out1;

%C1#3%

[IND1_T1 IND2_T1 IND3_T1] (m7-m9);

IND1_T1 IND2_T1 IND3_T1 (v7-v9);

[Out1] (oa3); Out1;

MODEL C2:

%C2#1%

[IND1_T2 IND2_T2 IND3_T2] (m1-m3);

IND1_T2 IND2_T2 IND3_T2 (v1-v3);

[Out2] (oa1); Out2;

%C2#2%

[IND1_T2 IND2_T2 IND3_T2] (m4-m6);

IND1_T2 IND2_T2 IND3_T2 (v4-v6);

```
[Out2] (oa2); Out2;
```

```
%C2#3%
```

```
[IND1_T2 IND2_T2 IND3_T2] (m7-m9);
```

```
IND1_T2 IND2_T2 IND3_T2 (v7-v9);
```

```
[Out2] (oa3); Out2;
```

MODEL CONSTRAINT:

```
NEW (y12); y12 = oa1-oa2;
```

```
NEW (y13); y13 = oa1-oa3;
```

```
NEW (y23); y23 = oa2-oa3;
```

LINEAR GROWTH MIXTURE ANALYSIS

In a latent curve model (LCM) estimated in *Mplus* the “I S |” command serves as a shortcut to define the latent intercept and slope factors, which is followed by a specification of the time-varying indicators and their time codes (i.e., λ_t ; their loadings on the intercept and slope factor).

In this input, we request the means of the intercepts and slope factors ([I S];), their variances (I S;), their covariances (I WITH S;), and all time specific residuals (IND1_T1, IND1_T2, IND1_T3, IND1_T4, IND1_T5, IND1_T6;) be freely estimated in all profiles. Here, we assume six repeated measures of a specific indicator (IND1_T1- IND1_T6; which would, of course, have been defined in the NAMES and USEVARIABLE commands), equally spaced (otherwise the slope loadings would need to be adjusted to reflect the actual time intervals), with an intercept located at Time 1. Code to plot these trajectories is provided on the following page.

```
%OVERALL%
```

```
I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;
```

```

%C#1%
[I S] (int1 slope1);
I S; I WITH S;
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%
[I S] (int2 slope2);
I S; I WITH S;
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%
[I S] (int3 slope3);
I S; I WITH S;
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

```

The following functions provide two alternative ways to plot these trajectories. The first way uses *Mplus*' default plotting procedure. The second function requires some additional user customization, but advantageously, can provide a plot with confidence intervals around the trajectories in each profile group, and is specified within a MODEL CONSTRAINT procedure. With the PLOT command, arbitrary labels are used to define the trajectories to be plotted. Conceivably there should be one for each profile group, and there are three in this example. The LOOP function then defines an arbitrary label for a variable that will be used in computing the equations for each trajectory. Here, "Time" is used to reflect the notion of growth curves. The rest of the LOOP function requests that "Time" is given an interval of 0 to 5, and is increased by 0.1 increments. So, in the computation of the following Profile1 trajectory, values

for “Time” between 0 and 5, increasing at 0.1 increments, will be included in the equation, and then the resulting values will be plotted. Here, the interval of 0 to 5 was chosen to reflect plotted the trajectories to start at same point as the measurement occasions (i.e., the intercept, before any linear change occurred), and six measures were obtained. These values can be adjusted for any particular application. Critically, this approach also requires the labeling of the mean intercept and slope mean parameters in each class (i.e., int1 and slope1) to be used in the calculations of the trajectories.

```
PLOT:
TYPE IS PLOT3;
SERIES = IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (*);
```

```
MODEL CONSTRAINT:
PLOT(Profile1 Profile2 Profile3);
LOOP(Time,0,5,0.1);
Profile1 = int1+Time*slope1;
Profile2 = int2+Time*slope2;
Profile3 = int3+Time*slope3;
```

As noted in our main description of linear growth mixture models, simpler parameterizations are possible should the above, fully variant model fail to converge on proper solutions. Our preference is to test these alternative models in the following order:

1. Alternative A – Homoscedastic Residuals (equal across time points)
2. Alternative B – Class-Invariant Residuals
3. Alternative C – Class-Invariant Latent Variance–Covariance Matrix
4. Alternative D – Class-Invariant Homoscedastic Residuals (equal across time points)

5. Alternative E – Class-Invariant Latent Variance–Covariance Matrix and Homoscedastic Residuals
6. Alternative – Class-Invariant Latent Variance–Covariance Matrix and Class-Invariant Residuals
7. Alternative G – Class-Invariant Latent Variance–Covariance Matrix and Class-Invariant Homoscedastic Residuals

Alternative A – Homoscedastic Residuals (equal across time points)

```

%OVERALL%
I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

%C#1%
[I S];
I S; I WITH S;
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (R1);

%C#2%
[I S];
I S; I WITH S;
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (R2);

%C#3%
[I S];
I S; I WITH S;
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (R3);

```

Alternative B – Class-Invariant Residuals

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

%C#1%

[I S];

I S; I WITH S;

%C#2%

[I S];

I S; I WITH S;

%C#3%

[I S];

I S; I WITH S;

Alternative C – Class-Invariant Latent Variance–Covariance Matrix

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

%C#1%

[I S];

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%

[I S];
 IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%

[I S];
 IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

Alternative D – Class-Invariant Homoscedastic Residuals (equal across time points)

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

%C#1%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (**R1**);

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (**R1**);

%C#3%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (**R1**);

Alternative E – Class-Invariant Latent Variance-Covariance Matrix and Homoscedastic Residuals

```
%OVERALL%
I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

%C#1%
[I S];
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (R1);

%C#2%
[I S];
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (R2);

%C#3%
[I S];
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (R3);
```

Alternative F – Class-Invariant Latent Variance-Covariance Matrix and Class-Invariant Residuals

Note that this is the default specification within *Mplus*.

```
%OVERALL%
I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

%C#1%
[I S];
```

```
%C#2%
```

```
[I S];
```

```
%C#3%
```

```
[I S];
```

Alternative G – Class-Invariant Latent Variance–Covariance Matrix and Class-Invariant Homoscedastic Residuals

```
%OVERALL%
```

```
I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;
```

```
%C#1%
```

```
[I S];
```

```
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (R1);
```

```
%C#2%
```

```
[I S];
```

```
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (R1);
```

```
%C#3%
```

```
[I S];
```

```
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (R1);
```

QUADRATIC GROWTH MIXTURE ANALYSIS

A quadratic slope factor (Q) can be added to allow for curvilinear trends. The following code provides the commands for the fully variant model, but the sequential testing described above can be applied to derive simpler parameterizations should issues of convergence or statistical adequacy result. The same two plotting procedures, noted above, can be applied, but we detail the LOOP approach below given the need for more user customization. If further polynomial functions are required, their factor label (i.e., CU for cubic [labeling it C would duplicate the label assigned to the latent profiles and would not be allowed], etc.) would just be added to the growth statement before the | symbol, similar to the difference between the linear and quadratic functions. Additional commands for the extra polynomial functions would then be added to each profile-specific section.

```
%OVERALL%

I S Q | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

%C#1%

I S Q;

[I S Q] (int1 linear1 quad1);

I WITH S Q;

S WITH Q;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%

I S Q;

[I S Q] (int2 linear2 quad2);

I WITH S Q;

S WITH Q;
```

```

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%

I S Q;

[I S Q] (int3 linear3 quad3);

I WITH S Q;

S WITH Q;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

MODEL CONSTRAINT:

PLOT(Profile1 Profile2 Profile3);

LOOP(Time,0,5,0.1);

Profile1 = int1+Time*linear1+quad1*Time**2;

Profile2 = int2+Time*linear2+quad2*Time**2;

Profile3 = int3+Time*linear3+quad3*Time**2;

```

PIECEWISE GROWTH MIXTURE ANALYSIS

A piecewise growth model implements multiple slope factors to account for distinct “before and after” trajectories that may be exhibited. Two linear slope factors (here, S1 and S2) can be defined to represent linear changes before and after a transition point. In this example, the second slope or trajectory begins at Time 4. As in previous models, simpler parameterizations are possible to implement should they be required. As well, piecewise models can be specified to include curvilinear trajectories before and/or after the transition point. For plotting, the LOOP function would not be readily available because *Mplus* can only allow for a single LOOP command in an analysis (it may however be possible to adjust the starting and ending points of the LOOP and import the calculated plot data from separate LOOPS into Microsoft Excel and replot the trajectories there, but this is not an overly straightforward procedure).

However, the built-in PLOT command can be used straightforwardly to visualize the functions.

```
%OVERALL%

I S1 | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@2 IND1_T5@2 IND1_T6@2;
I S2 | IND1_T1@0 IND1_T2@0 IND1_T3@0 IND1_T4@1 IND1_T5@2 IND1_T6@3;

%C#1%

I S1 S2;

[I S1 S2];

I WITH S1 S2;

S1 WITH S2;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%

I S1 S2;

[I S1 S2];

I WITH S1 S2;

S1 WITH S2;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%

I S1 S2;

[I S1 S2];

I WITH S1 S2;

S1 WITH S2;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

PLOT:

TYPE IS PLOT3;
```

```
SERIES = IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (*);
```

LATENT BASIS GROWTH MIXTURE ANALYSIS

In a latent basis model, two loadings (typically the first and last) need to be fixed, respectively, to 0 and 1 (i.e., using @0 and @1) while the other loadings are freely estimated (i.e., using *).

Here, we request that the loadings for the Time 2, 3, 4, and 5 assessments be freely estimated in all profiles by repeating these commands in the profile-specific sections. As in previous models, simpler parameterizations are possible to implement should they be required. The LOOP function will not adequately demonstrate the growth trajectories because the differing factor loadings on the slope factor (reflecting non-linearity) cannot be taken into account (i.e., the slope factor in a latent basis model only reflects the total change occurring over time). The PLOT command can therefore be used to visualize the growth trajectory

```
%OVERALL%
```

```
I S | IND1_T1@0 IND1_T2* IND1_T3* IND1_T4* IND1_T5* IND1_T6@1;
```

```
%C#1%
```

```
I S | IND1_T1@0 IND1_T2* IND1_T3* IND1_T4* IND1_T5* IND1_T6@1;
```

```
I S;
```

```
[I S] (int1 slope1);
```

```
I WITH S;
```

```
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;
```

```
%C#2%
```

```
I S | IND1_T1@0 IND1_T2* IND1_T3* IND1_T4* IND1_T5* IND1_T6@1;
```

```
I S;
```

```

[I S] (int2 slope2);
I WITH S;
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%
I S | IND1_T1@0 IND1_T2* IND1_T3* IND1_T4* IND1_T5* IND1_T6@1;
I S;
[I S] (int3 slope3);
I WITH S;
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

TYPE IS PLOT3:
SERIES = IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (*);

```

MEHTA AND WEST (2000) TEST

As change may occur differentially for cases that differ on tenure or age (at the first assessment), bias may be introduced into the meaning and interpretation of the latent slope. Mehta and West (2000) advised assessing the relation that age (or any other meaningful time-related variable) has on the estimate of the latent slope. In particular, they noted that variations in individuals' initial age/tenure/grade can be considered negligible when the following two conditions are met: (1) the regression of the intercept of a LCM on age is equal to the mean of the slope factor, and (2) the regression of the slope factor on age is equal to zero. Testing these conditions involves contrasting a typical, unrestricted LCM model that incorporates age as a predictor of the intercept and slope factors to a model in which these two conditions are imposed. If the more constrained model fits the data as well as the first model (as assessed via typical goodness-of-fit indices), then the impact of age can be considered to be negligible.

Initial Model

```

VARIABLE:

NAMES = ID Age IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

USEVARIABLES = Age IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

MISSING = all (999);

IDVARIABLE = ID;

ANALYSIS:

ESTIMATOR = MLR;

MODEL:

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

I ON Age;

S ON Age;

```

Constrained Model

```

MODEL:

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

I ON Age (c1);

[S] (c1);

S ON Age@0;

```

AGE-DEFINED LINEAR GROWTH MIXTURE ANALYSIS

To estimate a growth mixture analysis where age (or any other variable differing across individuals) is used to define the time codes (i.e., factor loadings for the latent slope), this variable needs to be identified using the CONSTRAINT command of the VARIABLE section

of the syntax. The **CONSTRAINT** command identifies variables used only in the **MODEL CONSTRAINT** section of the syntax. In the age-defined growth model, the loadings on the slope factor (representing the time code), rather than being fixed to specific values, are freely estimated and assigned a label (here, L1 to L6). Using this approach involves relying on the complete specification of the LCM model (without being able to use the shortcut function associated with the | symbol). To apply a complete specification, one also has to constrain the time-specific intercepts to 0, and request the free estimation of the intercept and slope factor means in the **%OVERALL%** section. In this example, we define trajectories on the basis of individually-varying ages (coded in years), and want to set the intercept of the trajectories to that of a 25-year-old (hence the subtraction of 25 in each of the **MODEL CONSTRAINT** equations; this corresponds to the age of the youngest employee at Time 1). Measurement occurred at annual intervals that were equal for all participants. If measurement intervals also differed across participants, then one **AGE** variable would be required for each time of measurement (i.e., **Age_T1**, **Age_T2**, etc.). With this approach, time-specific residuals should be specified as homoscedastic (equal across time points) as the alternative approach of allowing them to differ across time of measurement would be inconsistent with the definition of time codes that ignore the time of measurement. The default **PLOT** function does not work in this context, so that the trajectories would need to be plotted with the **LOOP** function described earlier, or graphed manually. This model is illustrated below.

```
VARIABLE:  
  
NAMES = ID Age IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;  
  
USEVARIABLES = IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;  
  
MISSING = all (999);  
  
IDVARIABLE = ID;  
  
CONSTRAINT = Age;
```

```

CLASSES = C(3);

ANALYSIS:

TYPE = MIXTURE COMPLEX; ESTIMATOR = MLR;

PROCESS = 3; STARTS = 3000 100; STITERATIONS = 100;

MODEL:

%OVERALL%

I BY IND1_T1@1 IND1_T2@1 IND1_T3@1 IND1_T4@1 IND1_T5@1 IND1_T6@1;

S BY IND1_T1* (L1)

    IND1_T2 (L2)

    IND1_T3 (L3)

    IND1_T4 (L4)

    IND1_T5 (L5)

    IND1_T6 (L6);

I S; I WITH S; [I S];

[IND1_T1@0 IND1_T2@0 IND1_T3@0 IND1_T4@0 IND1_T5@0 IND1_T6@0];

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#1%

I S; I WITH S;

[I S] (int1 slope1);

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (r1);

%C#2%

I S; I WITH S;

[I S] (int2 slope2);

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (r2);

```

```

%C#3%
I S; I WITH S;
[I S] (int3 slope3);
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (r3);

MODEL CONSTRAINT:

L1 = AGE - 25;
L2 = AGE - 25 + 1;
L3 = AGE - 25 + 2;
L4 = AGE - 25 + 3;
L5 = AGE - 25 + 4;
L6 = AGE - 25 + 5;

MODEL CONSTRAINT:
PLOT(Profile1 Profile2 Profile3); LOOP(Time,0,5,0.1);
Profile1 = int1+Time*slope1;
Profile2 = int2+Time*slope2;
Profile3 = int3+Time*slope3;

```

AGE-DEFINED QUADRATIC GROWTH MIXTURE ANALYSIS

A similar approach can be used for a quadratic model (or any other polynomial models, where the form of the polynomial would apply to the equations used within the MODEL CONSTRAINT; i.e., the functions are squared in the below quadratic formulation). As above, the LOOP function or manual plotting of the resulting trajectories would be needed.

MODEL:

%OVERALL%

I BY IND1_T1@1 IND1_T2@1 IND1_T3@1 IND1_T4@1 IND1_T5@1 IND1_T6@1;

S BY IND1_T1* (L1)

IND1_T2 (L2)

IND1_T3 (L3)

IND1_T4 (L4)

IND1_T5 (L5)

IND1_T6 (L6);

Q BY IND1_T1* (Q1)

IND1_T2 (Q2)

IND1_T3 (Q3)

IND1_T4 (Q4)

IND1_T5 (Q5)

IND1_T6 (Q6);

I S Q; I WITH S Q; S WITH Q;

[I S Q];

[IND1_T1@0 IND1_T2@0 IND1_T3@0 IND1_T4@0 IND1_T5@0 IND1_T6@0];

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#1%

I S Q; I WITH S Q; S WITH Q;

[I S Q] (int1 linear1 quad1);

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (**r1**);

%C#2%

I S Q; I WITH S Q; S WITH Q;

```
[I S Q] (int2 linear2 quad2);
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (r2);
%C#3%
I S Q; I WITH S Q; S WITH Q;
[I S Q] (int3 linear3 quad3);
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6 (r3);

MODEL CONSTRAINT:

L1 = AGE - 25;
L2 = AGE - 25 + 1;
L3 = AGE - 25 + 2;
L4 = AGE - 25 + 3;
L5 = AGE - 25 + 4;
L6 = AGE - 25 + 5;
Q1 = (AGE - 25)**2;
Q2 = (AGE - 25+1)**2;
Q3 = (AGE - 25+2)**2;
Q4 = (AGE - 25+3)**2;
Q5 = (AGE - 25+4)**2;
Q6 = (AGE - 25+5)**2;

MODEL CONSTRAINT:
PLOT(Profile1 Profile2 Profile3);
LOOP(Time,0,5,0.1);
Profile1 = int1+Time*linear1+quad1*Time**2;
```

```
Profile2 = int2+Time*linear2+quad2*Time**2;
```

```
Profile3 = int3+Time*linear3+quad3*Time**2;
```

GROWTH MIXTURE ANALYSIS: TIME-INVARIANT PREDICTORS

The following models involve contrasting a model in which the relations for the time-invariant predictors are restricted to zero (i.e., the null effects model), and alternative models in which (a) the predictors may only potentially influence class membership, (b) the predictors can influence class membership, and also the intercept and slope factors, but that the intercept and slope relations are held equal across profile groups, and (c) the predictors can influence class membership, and can also influence the intercept and slope freely across profiles. Adjustments to the syntax to facilitate these different models are highlighted in bold below.

Null-Effects Model

```
%OVERALL%
```

```
I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;
```

```
C ON Pred1@0 Pred2@0;
```

```
I ON Pred1@0 Pred2@0;
```

```
S ON Pred1@0 Pred2@0;
```

```
%C#1%
```

```
[I S];
```

```
I S; I WITH S;
```

```
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;
```

```
%C#2%
```

```
[I S];
```

I S; I WITH S;
 IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%

[I S];
 I S; I WITH S;
 IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

Alternative A – Freely Estimated Effects on Class Membership Only

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

C ON Pred1 Pred2;

I ON Pred1@0 Pred2@0;

S ON Pred1@0 Pred2@0;

%C#1%

[I S];
 I S; I WITH S;
 IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%

[I S];
 I S; I WITH S;
 IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

Alternative B – Effects on Class Membership and Class-Invariant Effects on the Growth

Factors

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

C ON Pred1 Pred2;

I ON Pred1 Pred2;

S ON Pred1 Pred2;

%C#1%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%

[I S];

I S; I WITH S;
 IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

Alternative C – Effects on Class Membership and Class-Varying Effects on the Growth Factors

%OVERALL%
 I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;
 C ON Pred1 Pred2;
 I ON Pred1 Pred2;
 S ON Pred1 Pred2;

%C#1%
 [I S];
 I S; I WITH S;
 IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;
I ON Pred1 Pred2;
S ON Pred1 Pred2;

%C#2%
 [I S];
 I S; I WITH S;
 IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;
I ON Pred1 Pred2;
S ON Pred1 Pred2;

```

%C#3%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

I ON Pred1 Pred2;

S ON Pred1 Pred2;

```

GROWTH MIXTURE ANALYSIS: TIME-INVARIANT OUTCOMES

These analyses follow from the LPA with outcomes models described earlier.

```

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

%C#1%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

[Out1] (oa1);

[Out2] (ob1);

Out1;

Out2;

%C#2%

[I S];

I S; I WITH S;

```

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

[Out1] (oa2);

[Out2] (ob2);

Out1;

Out2;

%C#3%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

[Out1] (oa3);

[Out2] (ob3);

Out1;

Out2;

MODEL CONSTRAINT:

NEW (y12);

y12 = oa1-oa2;

NEW (y13);

y13 = oa1-oa3;

NEW (y23);

y23 = oa2-oa3;

NEW (z12);

z12 = ob1-ob2

NEW (z13);

```
z13 = ob1-ob3;
```

```
NEW (z23);
```

```
z23 = ob2-ob3;
```

GROWTH MIXTURE ANALYSIS: TIME-VARYING PREDICTORS

The following series of models presents a potential framework that can be used for time-varying predictors. In contrast to the previous predictor models, where the predictors targeted profile membership, the intercept and slope factors, or both, time-varying predictors target the focal indicators that have been measured at each occasion, respectively. As above, these can also involve a contrast between a model in which the relations for the time-varying predictors are restricted to zero (i.e., the null effects model), and alternative models that include (a) effects that are free to vary across time points, but equal across profiles, (b) effects that are free to vary across time points and profiles, (c) effects that are free to vary across profiles but not time points, and (d) effects constrained to equality across time points and profiles.

Null-Effects Model

```
%OVERALL%
```

```
I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;
```

```
IND1_T1 ON PredT1@0;
```

```
IND1_T2 ON PredT2@0;
```

```
IND1_T3 ON PredT3@0;
```

```
IND1_T4 ON PredT4@0;
```

```
IND1_T5 ON PredT5@0;
```

```
IND1_T6 ON PredT6@0;
```

```
[PredT1] (m1); [PredT2] (m2); [PredT3] (m3); [PredT4] (m4); [PredT5] (m5); [PredT6] (m6);
```

%C#1%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

Alternative A – Effects Free to Vary Across Time Points but Equal Across Profiles

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

IND1_T1 ON PredT1;

IND1_T2 ON PredT2;

IND1_T3 ON PredT3;

IND1_T4 ON PredT4;

IND1_T5 ON PredT5;

IND1_T6 ON PredT6;

[PredT1] (m1); [PredT2] (m2); [PredT3] (m3); [PredT4] (m4); [PredT5] (m5); [PredT6] (m6);

%C#1%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

Alternative B – Effects Free to Vary Across Time Points and Profiles

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

IND1_T1 ON PredT1;

IND1_T2 ON PredT2;

IND1_T3 ON PredT3;

IND1_T4 ON PredT4;

IND1_T5 ON PredT5;

IND1_T6 ON PredT6;

[PredT1] (m1); [PredT2] (m2); [PredT3] (m3); [PredT4] (m4); [PredT5] (m5); [PredT6] (m6);

%C#1%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

IND1_T1 ON PredT1;

IND1_T2 ON PredT2;

IND1_T3 ON PredT3;

IND1_T4 ON PredT4;

IND1_T5 ON PredT5;

IND1_T6 ON PredT6;

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

IND1_T1 ON PredT1;

IND1_T2 ON PredT2;

IND1_T3 ON PredT3;

IND1_T4 ON PredT4;

IND1_T5 ON PredT5;

IND1_T6 ON PredT6;

%C#3%

```

[I S];
I S; I WITH S;
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;
IND1_T1 ON PredT1;
IND1_T2 ON PredT2;
IND1_T3 ON PredT3;
IND1_T4 ON PredT4;
IND1_T5 ON PredT5;
IND1_T6 ON PredT6;

```

Alternative C – Effects Free to Vary Across Profiles but Not Time Points

```

%OVERALL%
I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;
IND1_T1 ON PredT1;
IND1_T2 ON PredT2;
IND1_T3 ON PredT3;
IND1_T4 ON PredT4;
IND1_T5 ON PredT5;
IND1_T6 ON PredT6;
[PredT1] (m1); [PredT2] (m2); [PredT3] (m3); [PredT4] (m4); [PredT5] (m5); [PredT6] (m6);

%C#1%
[I S];
I S; I WITH S;
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

```

IND1_T1 ON PredT1 (**r1**);

IND1_T2 ON PredT2 (**r1**);

IND1_T3 ON PredT3 (**r1**);

IND1_T4 ON PredT4 (**r1**);

IND1_T5 ON PredT5 (**r1**);

IND1_T6 ON PredT6 (**r1**);

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

IND1_T1 ON PredT1 (**r2**);

IND1_T2 ON PredT2 (**r2**);

IND1_T3 ON PredT3 (**r2**);

IND1_T4 ON PredT4 (**r2**);

IND1_T5 ON PredT5 (**r2**);

IND1_T6 ON PredT6 (**r2**);

%C#3%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

IND1_T1 ON PredT1 (**r3**);

IND1_T2 ON PredT2 (**r3**);

IND1_T3 ON PredT3 (**r3**);

IND1_T4 ON PredT4 (**r3**);

IND1_T5 ON PredT5 (**r3**);

IND1_T6 ON PredT6 (**r3**);

Alternative D – Effects Constrained to Equality Across Time Points and Profiles

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

IND1_T1 ON PredT1 (**r1**);

IND1_T2 ON PredT2 (**r1**);

IND1_T3 ON PredT3 (**r1**);

IND1_T4 ON PredT4 (**r1**);

IND1_T5 ON PredT5 (**r1**);

IND1_T6 ON PredT6 (**r1**);

[PredT1] (m1); [PredT2] (m2); [PredT3] (m3); [PredT4] (m4); [PredT5] (m5); [PredT6] (m6);

%C#1%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

```
%C#3%
```

```
[I S];
```

```
I S; I WITH S;
```

```
IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;
```

GROWTH MIXTURE ANALYSIS: TIME-VARYING OUTCOMES

The final series of models we present can be leveraged to investigate time-varying outcome relations. These are similar to the previous set of models that specified time-varying predictors. In contrast to the models where time-invariant outcomes were specified, these outcome relations are specified as a function of the focal indicators that have been measured at each occasion. Further, in this case the main interest might be to examine the regression relations between the time-specific indicators and the time-specific outcomes, rather than mean differences in the outcomes across the mixture groups (though they are still estimated, and may of course be of substantive interest). As in the time-varying predictor models, determining an appropriate, optimal model can involve contrasts between a model in which the relations for the time-varying predictors are restricted to zero (i.e., the null effects model), and alternative models that include (a) effects that are free to vary across time points, but equal across profiles, (b) effects that are free to vary across time points and profiles, (c) effects that are free to vary across profiles but not time points, and (d) effects constrained to equality across time points and profiles. In the following examples, we (a) freely estimate the full regression equation (which includes the slope, intercepts, and residuals) associated with the time-varying outcomes and allow them to differ, and (b) constrain these three components (slope, intercepts, and residuals) to equality across profiles and time points. Simpler models only specifying differences in regression slopes can also be estimated, in which case the syntax presented in greyscale should be omitted. Alternatively, equality constraints can be imposed and tested in

sequence (on the slopes, then intercepts, then residuals) once the optimal model has been selected.

Null-Effects Model

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

OutT1 ON IND1_T1@0;

OutT2 ON IND1_T2@0;

OutT3 ON IND1_T3@0;

OutT4 ON IND1_T4@0;

OutT5 ON IND1_T5@0;

OutT6 ON IND1_T6@0;

[OutT1] (m1); [OutT2] (m2); [OutT3] (m3); [OutT4] (m4); [OutT5] (m5); [OutT6] (m6);

%C#1%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

Alternative A – Effects Free to Vary Across Time Points but Equal Across Profiles

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

OutT1 ON IND1_T1;

OutT2 ON IND1_T2;

OutT3 ON IND1_T3;

OutT4 ON IND1_T4;

OutT5 ON IND1_T5;

OutT6 ON IND1_T6;

[OutT1] (m1); [OutT2] (m2); [OutT3] (m3); [OutT4] (m4); [OutT5] (m5); [OutT6] (m6);

%C#1%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%

[I S];
 I S; I WITH S;
 IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

Alternative B – Effects Free to Vary Across Time Points and Profiles

%OVERALL%
 I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;
 OutT1 ON IND1_T1;
 OutT2 ON IND1_T2;
 OutT3 ON IND1_T3;
 OutT4 ON IND1_T4;
 OutT5 ON IND1_T5;
 OutT6 ON IND1_T6;

%C#1%
 [I S];
 I S; I WITH S;
 IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;
OutT1 ON IND1_T1;
OutT2 ON IND1_T2;
OutT3 ON IND1_T3;
OutT4 ON IND1_T4;
OutT5 ON IND1_T5;
OutT6 ON IND1_T6;
 [OutT1] (m1); [OutT2] (m2); [OutT3] (m3); [OutT4] (m4); [OutT5] (m5); [OutT6] (m6);

OutT1 (v1); OutT2 (v2); OutT3 (v3); OutT4 (v4); OutT5 (v5); OutT6 (v6);

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

OutT1 ON IND1_T1;

OutT2 ON IND1_T2;

OutT3 ON IND1_T3;

OutT4 ON IND1_T4;

OutT5 ON IND1_T5;

OutT6 ON IND1_T6;

[OutT1] (n1); [OutT2] (n2); [OutT3] (n3); [OutT4] (n4); [OutT5] (n5); [OutT6] (n6);

OutT1 (vv1); OutT2 (vv2); OutT3 (vv3); OutT4 (vv4); OutT5 (vv5); OutT6 (vv6);

%C#3%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

OutT1 ON IND1_T1;

OutT2 ON IND1_T2;

OutT3 ON IND1_T3;

OutT4 ON IND1_T4;

OutT5 ON IND1_T5;

OutT6 ON IND1_T6;

[OutT1] (p1); [OutT2] (p2); [OutT3] (p3); [OutT4] (p4); [OutT5] (p5); [OutT6] (p6);
 OutT1 (vvv1); OutT2 (vvv2); OutT3 (vvv3); OutT4 (vvv4); OutT5 (vvv5); OutT6 (vvv6);

Alternative C – Effects Free to Vary Across Profiles but Not Time Points

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

OutT1 ON IND1_T1;

OutT2 ON IND1_T2;

OutT3 ON IND1_T3;

OutT4 ON IND1_T4;

OutT5 ON IND1_T5;

OutT6 ON IND1_T6;

%C#1%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

OutT1 ON IND1_T1 (o1);

OutT2 ON IND1_T2 (o1);

OutT3 ON IND1_T3 (o1);

OutT4 ON IND1_T4 (o1);

OutT5 ON IND1_T5 (o1);

OutT6 ON IND1_T6 (o1);

[OutT1] (m1); [OutT2] (m1); [OutT3] (m1); [OutT4] (m1); [OutT5] (m1); [OutT6] (m1);

OutT1 (v1); OutT2 (v1); OutT3 (v1); OutT4 (v1); OutT5 (v1); OutT6 (v1);

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

OutT1 ON IND1_T1 (o2);

OutT2 ON IND1_T2 (o2);

OutT3 ON IND1_T3 (o2);

OutT4 ON IND1_T4 (o2);

OutT5 ON IND1_T5 (o2);

OutT6 ON IND1_T6 (o2);

[OutT1] (m2); [OutT2] (m2); [OutT3] (m2); [OutT4] (m2); [OutT5] (m2); [OutT6] (m2);

OutT1 (v2); OutT2 (v2); OutT3 (v2); OutT4 (v2); OutT5 (v2); OutT6 (v2);

%C#3%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

OutT1 ON IND1_T1 (o3);

OutT2 ON IND1_T2 (o3);

OutT3 ON IND1_T3 (o3);

OutT4 ON IND1_T4 (o3);

OutT5 ON IND1_T5 (o3);

OutT6 ON IND1_T6 (o3);

[OutT1] (m3); [OutT2] (m3); [OutT3] (m3); [OutT4] (m3); [OutT5] (m3); [OutT6] (m3);

OutT1 (v3); OutT2 (v3); OutT3 (v3); OutT4 (v3); OutT5 (v3); OutT6 (v3);

Alternative D – Effects Constrained to Equality Across Time Points and Profiles

%OVERALL%

I S | IND1_T1@0 IND1_T2@1 IND1_T3@2 IND1_T4@3 IND1_T5@4 IND1_T6@5;

OutT1 ON IND1_T1 (o1);

OutT2 ON IND1_T2 (o1);

OutT3 ON IND1_T3 (o1);

OutT4 ON IND1_T4 (o1);

OutT5 ON IND1_T5 (o1);

OutT6 ON IND1_T6 (o1);

[OutT1] (m1); [OutT2] (m1); [OutT3] (m1); [OutT4] (m1); [OutT5] (m1); [OutT6] (m1);

OutT1 (v1); OutT2 (v1); OutT3 (v1); OutT4 (v1); OutT5 (v1); OutT6 (v1);

%C#1%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#2%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;

%C#3%

[I S];

I S; I WITH S;

IND1_T1 IND1_T2 IND1_T3 IND1_T4 IND1_T5 IND1_T6;