Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: www.tandfonline.com/journals/hsem20

Methodological Advances with Penalized Structural Equation Models

Tihomir Asparouhov & Bengt Muthén

STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
2025, VOL. 32, NO. 4, 688–716
https://doi.org/10.1080/10705511.2024.2425996

Routledge
Taylor & Francis Group

TEACHER'S CORNER

Methodological Advances with Penalized Structural Equation Models

Tihomir Asparouhov and Bengt Muthén

Mplus

ABSTRACT
Penalized structural equation models (PSEM) is a powerful technique that unlocks a variety of new modeling frameworks. PSEM applications have been established previously for standard SEM and ESEM models. In this note we aim to extend these ideas to more general types of models such as finite mixture models, multilevel models as well as models with more general types of outcomes. Maximum likelihood and weighted least squares estimation methods naturally accommodate a penalty term. In Mplus 8.12 the PSEM methodology is implemented for all models that can be estimated with these two estimators. Therefore we can now easily combine the more general models with the features of PSEM such as EFA, Alignment, and parameter invariance. Some additional basic SEM applications are also included.

KEYWORDS
Alignment; EFA; penalized maximum-likelihood; regularization

1. Introduction

The penalized structural equation models (PSEM) described in Asparouhov and Muthén (2024) have been used so far within the SEM and ESEM modeling frameworks to create new models that can address practical challenges which standard modeling methods can not address. The PSEM methodology is available for maximum-likelihood (ML) and weighted least squares (WLS) estimation in Mplus under the names PML and PWLS. In Mplus 8.12 the PML and PWLS estimators are extended to all models that can be estimated in Mplus with the ML and WLS estimators. This means that new modeling opportunities are now available for the PSEM framework. In particular, finite mixture and multilevel models can now be estimated with predesigned penalty to incorporate PSEM features such as factor rotation, factor analysis alignment, as well as parameter invariance across structures in these more general settings. More specifically, the PSEM extensions included in Mplus 8.12 are as follows.

- Maximum likelihood estimation of single level and two-level SEM models with numerical integration which includes a variety of dependent variable types such as continuous, categorical, censored, count, nominal, and survival variables.
- Maximum likelihood EM-algorithm based estimation of two-level and three-level SEM models with continuous dependent variables.
- Maximum likelihood EM-algorithm based estimation of single level finite mixture models.
- Maximum likelihood numerical integration based estimation of single and two-level finite mixture models.
- Weighted least squares estimation of two-level SEM models with continuous and categorical dependent variables.

In this article we aim to illustrate new models that are now available in these settings. In addition, some new basic PSEM concepts will be illustrated as well. The article is intended to be an inspiration for real-data applications. We illustrate general concepts and methodological advances. How exactly these new possibilities will be connected to real data applications is still a wide open and largely unexplored area of research.

2. Summary of the PSEM Framework

The PSEM framework in Mplus is based on adding a penalty function to an existing estimation method for the purpose of obtaining a very specific model estimation. Two estimators have been utilized for this framework so far: ML (maximum-likelihood) and WLS (weighted least squares). Both of these estimators minimize a fit function $F(\theta)$, where θ is the vector of model parameters. In the PSEM framework we add a weighted penalty to the fit function and the estimator now minimizes

$$F_1(\theta) = F(\theta) + wP(\theta) \quad (1)$$

where w is the penalty weight and P is the penalty function.

Two different types of models are estimated using this framework. The first type is a PSEM model. This is an unidentified model, i.e., if the model is estimated by minimizing only the fit function $F(\theta)$ using the standard ML or WLS estimators, the model becomes unidentified: an entire subspace of the θ parameters yield the same $F(\theta)$ values. The model becomes identified when the penalty is added to

CONTACT Tihomir Asparouhov tihomir@statmodel.com Research and Development, Muthen & Muthen, Los Angeles, CA, USA.

the fit function. In this estimation, the weight w is set to a small value so that when we optimize $F_1(\theta)$ we primarily optimize $F(\theta)$. More specifically, when w is sufficiently small, optimizing $F_1(\theta)$ is equivalent to optimizing $F(\theta)$ first, and in the subspace of θ where $F(\theta)$ is maximized, the penalty $P(\theta)$ is then optimized. This secondary optimization leads to identifying the model. A PSEM model estimation may also describe a null model: this is a model in the subspace where $F(\theta)$ is maximized and is a known model. The null model is identified by explicitly constraining some of the θ parameters. By comparing the fit function values $F(\theta)$ for the PSEM and the null model, we can ensure that the two models yield the same data fit. The weight w is chosen numerically as the largest value yielding the same data fit for the PSEM and the null model.

The second type of models that are estimated with the PSEM framework are the Regularized SEM models (RegSEM). The RegSEM model is identified even if we use the standard estimation without the penalty. For the RegSEM estimation, when we add the penalty to the fit function, a portion of the $F(\theta)$ optimization is sacrificed to the benefit of optimizing the penalty $P(\theta)$. The weight w is selected so that it is not too big to damage substantially the data fit optimization of $F(\theta)$ and not too small so that the penalty effect is not eliminated.

In Asparouhov and Muthén (2024) it was shown that EFA models are a special case of PSEM models where the penalty is set to the rotation criterion. It was shown also that the multiple group alignment method of Asparouhov and Muthén (2014b) is a special case of PSEM models where the penalty function is set to the alignment optimization function. It was also shown that the BSEM methodology of Muthén and Asparouhov (2012), which develops the concepts of approximately zero parameters and approximately equal parameters, is also a special case of PSEM models. These facts allow us to include optimal factor rotations, parameter alignment, approximate parameter equality, and approximately zero parameters in any general model.

In Mplus, the penalty function is specified with MODEL PRIOR. Any kind of penalty function can be specified but the most common univariate and multivariate penalty functions can be specified directly as priors. Three univariate penalties are available directly: θ^2, $|\theta|$, and $\sqrt{|\theta|}$ and these are specified as univariate priors for the parameters: $N(0, v)$, $LASSO(0, v)$, and $ALF(0, v)$ where $w = 1/v$ is the penalty weight. These are used when a parameters is meant to be approximately zero. Two multivariate penalties are also directly available. The first uses the *DIFF* function. For example, $DIFF(\theta_1 - \theta_k) \sim LASSO(0, v)$ results in a penalty function which is the sum of all absolute value pairwise differences in the set of parameters $\theta_1, \ldots, \theta_k$. The weight is again $w = 1/v$. The *DIFF* penalty/prior is used to ensure approximate equality in the set of parameters. The second multivariate penalty that is directly available is the Geomin rotation criterion. The specification $\theta_1 - \theta_k \sim GEOMIN(m, v)$ gives the Geomin rotation function for a loading matrix of size m by k/m with all the loadings parameters listed column by column as $\theta_1, \ldots, \theta_k$. This penalty is used to conduct EFA analysis and to rotate the loading matrix to the simplest pattern.

3. Algorithms for Maximizing the Penalized Fit Function

The scope of the PSEM generalization we describe here involves a variety of advanced numerical algorithms: numerical integration for continuous latent variables, EM-algorithm for categorical latent variables, EM-algorithm for missing data, EM-algorithm for random intercepts and slopes, acceleration for EM-algorithms. These advanced techniques are essential in the corresponding modeling frameworks. Because of these techniques we are able to efficiently estimate increasingly large and sophisticated latent variable models. It is therefore imperative to be able to incorporate the penalized likelihood into these methods. Fortunately this is a trivial task and it only depends on the approach.

There are several different approaches that can be used here. The first approach is to use an optimization method which is based only on the first or the first and the second derivatives of the function we are optimizing. To implement this approach the first and the second derivatives of the penalty are simply added to the first and the second derivatives of the data fit function. Algorithms that fall into this category are Quasi-Newton, Fisher-scoring, Newton-Raphson, and when these are used as accelerators to the EM-algorithm.

The second approach is based on the observation established in Asparouhov and Muthén (2024) that the PSEM model estimation is equivalent to the standard fit function optimization with parameter constraints. These constraints can be solved implicitly or explicitly and the log-likelihood constrained optimization is performed as usual. None of the EM-algorithms or the numerical integration is entangled with the constraints on the parameters. Mplus algorithms have been developed to use parameter constraints of any kind and therefore the PSEM optimization can take advantage of that prior development.

The third approach is based on treating the penalty as the log-likelihood of one additional observation that has its own model and is in its own group. This penalty model does not contain any latent variables and therefore when latent variables are dealt with using the EM-algorithm or numerical integration, the penalty will have no impact. The E-step of the EM-algorithm will remain unchanged. The M-step however will incorporate the penalty in the usual way as an addition. Since the M-step is derivative based, the incorporation of the penalty in that step is also easy.

For finite mixture models, the Mplus algorithm also takes advantage of separating the variables by indicator type: continuous vs. categorical vs. counts, etc. This way a large optimization problem is broken up into several smaller optimization problems. The smaller optimization problems where variables are of the same type may also allow special simplifications based on sufficient statistics, which would not be the case without the split. This M-step split is also acceptable for the penalized likelihood since the penalty usually splits nicely by variable type as well. For example, the Geomin rotation criterion is the sum of penalties for each row in the loading matrix, i.e., the penalty is variable specific and can be separated by variable types as well. In the M-step split, each of the smaller optimizations will include a portion of the penalty.

In the following sections, we describe a variety of new PSEM application areas.

4. Formative Factor

Consider the following factor analysis model. Let Y_p, $p = 1,\ldots,P$ be a set of factor indicators and X_q, $q = 1,\ldots,Q$, be a set of factor predictors. Consider the following model

$$Y_p = \nu_p + \lambda_p F + \varepsilon_p \tag{2}$$

$$F_0 = \sum_{q=1}^{Q} \beta_q X_q \tag{3}$$

$$F = \gamma F_0 + \xi \tag{4}$$

$$\xi \sim N(0, \phi), \varepsilon_p \sim N(0, \theta_p). \tag{5}$$

For identification purposes λ_1 and β_1 are fixed to 1. In the above model, the factor F is a reflective factor, while the factor F_0 is a formative factor, see Bollen (1989). The formative factor does not have a residual and is simply a weighted sum of the predictors. We can use the standard chi-square test to evaluate the model. However, the test does not address the existence of the formative factor. The test of fit can be rejected because of 3 different issues: covariance between the indicators is not explained by the reflective factor, there are direct effects from the formative factor to the indicators, or there is a direct effect from the covariates to the indicators.

The above model is equivalent to the MIMIC model where the factor F is regressed directly on all the covariates and the formative factor is excluded from the model. The formative factor becomes more valuable when more than one variable is regressed on that factor. In that case, the existence of the formative factor is essentially a hypothesis for regression coefficient proportionality if the raw covariates are used directly as predictors instead of via a formative factor. The general idea behind a formative factor is to replace a larger number of covariates by the most suitable linear combination of the covariates, which increases the power of the model. For example, variables such as education level, occupation, income and other background variables can be clumped together into a single predictor usually named SES (socioeconomic status). A different example is the case where rare event covariates yield weak power but when properly combined can provide a single meaningful predictor.

Here we provide a PSEM based factor analysis model that can be used more directly to support the existence of the formative factor. Consider the following PSEM model

$$Y_p = \nu_p + \lambda_p F + a_p F_0 + \sum_{q=1}^{Q} b_{pq} X_q + \varepsilon_p \tag{6}$$

$$F_0 = \sum_{q=1}^{Q} \beta_q X_q \tag{7}$$

$$F = \gamma F_0 + \xi \tag{8}$$

$$\xi \sim N(0, \phi), \varepsilon_p \sim N(0, \theta_p) \tag{9}$$

$$a_p \sim ALF(0,1), b_{pq} \sim ALF(0,1) \tag{10}$$

The parameters a_p account for possible direct effects from the formative factor to the indicators, while b_{pq} provide direct effects from the covariates to the indicators. If the parameters b_{pq} are all near zero and not significant, we interpret this as evidence for the existence of the formative factor. If a small portion of these parameters are significant, that means that in addition to the formative factor a small set of the covariates must be retained alongside the formative factor.

Because of the added penalties for these new parameters a_p and b_{pq}, we expect that the PSEM estimation will set these to zeros as long as the model fit is not compromised. If model (2–5) does not fit the data, some of the parameters a_p and b_{pq} will be non-zero, while the parameters β_q and λ_p will remain unbiased. Most importantly, the above PSEM model allows us to test the hypothesis that a linear combination of the covariates is sufficient to capture the predictive power of these covariates without entangling that hypothesis with the hypothesis that all the effects of the covariates go through the factor F.

We illustrate the PSEM model with a simulation study using 100 replications and sample size of 2000. The model has 4 indicator variable, 5 covariates, one direct effect from a covariate (one non-zero b_{pq} parameter) and one direct effect from the formative factor (one non-zero a_p parameter). Figure 1 shows the Mplus input file for this simulation study and Figure 2 shows the results for a selection of the parameters. The bias in the estimate is minimal and the coverage is near the nominal levels. For this particular model the formative factor can be retained using the estimated weights, while in addition to the formative factor, one more covariate X_3 must be retained due to its predictive power for indicator Y_2 that is significant and substantial. The chi-square test for this PSEM model has an average value of 2.2 and with 2 degrees of freedom the test yields a rejection rate of 8% which is near the nominal level. The MIMIC model, without the formative factor, yields a rejection rate of 100% and it will require the addition of 6 indirect effects to be acceptable (5 indirect effects stemming from the one indirect effect of the formative factor plus 1 indirect effect from X_3). The MIMIC model would also fail to recognize the uniform impact of the covariates as provided by the formative factor.

In Figure 3, we also provide the model statement for this formative factor analysis that can be used with real data sets. We provide this statement as it is much simpler than the model statement used in simulation study. A simulation study setup needs starting values to compute confidence intervals coverage.

```
MONTECARLO:
    NAMES = y1-y4 x1-x5;
    NOBSERVATIONS = 2000;
    NREPS = 100;

MODEL POPULATION:
    f1 BY y1*1 y2*0.8 y3*0.4 y4*0.7; f1*1 y1-y4*1;
    f2 by;
    f2 ON x1*1 x2*0.2 x3*1.1 x4*0.3 x5*0.4; f2@0;
    y3 on f2*0.2;
    x1-x5*1;
    f1 on f2*0.5;
    y2 on x3*0.4;

MODEL:
    f1 BY y1@1 y2*0.8 y3*0.4 y4*0.7; f1*1 y1-y4*1;
    f2 by;
    f2 ON x1@1 x2*0.2 x3*1.1 x4*0.3 x5*0.4; f2@0;
    y1 on f2*0 (a1);
    y2 on f2*0 (a2);
    y3 on f2*0.2 (a3);
    y4 on f2*0 (a4);
    f1 on f2*0.5;
    y1 on x1-x5*0 (b1-b5);
    y2 on x1-x2*0 x3*0.4 x4-x5*0 (b6-b10);
    y3 on x1-x5*0 (b11-b15);

MODEL PRIORS:
    a1-a4~ALF(0,1);
    b1-b15~ALF(0,1);
```

Figure 1. Formative factor simulation study.

5. Longitudinal Growth Modeling for Exploratory Factor Analysis

In this section we illustrate how PSEM can be used for longitudinal growth modeling of factors repeatedly measured by EFA. Such models can be used to study developmental changes with age across various traits that are measured via exploratory analysis. For example the big 5 personality traits are measured most accurately within the ESEM framework, see Marsh et al. (2010). Suppose that Y_{tp} are P factor measurements, $p = 1,..,P$ across various time points $t = 0,...,T$. Let M be the number of factors measured by these indicators and f_{tm} be the $m-$th factor measured at time point t, $m = 1,...,M$ and $t = 0,...,T$. The EFA measurement model at time t is given by the following equation

$$Y_{tp} = \nu_p + \sum_{m=1}^{M} \lambda_{pm}\eta_{tm} + \varepsilon_{tp} \quad (11)$$

The growth model for these factors is given by

$$\eta_{tm} = I_m + tS_m + \xi_{tm}, \quad (12)$$

where I_m is the latent intercept (the time invariant part of the $m-$th factor) and S_m is the latent slope factor (the systematic change across time in the $m-th$ factor). The variance covariance for the $2M$ latent variable I_m and S_m is unconstrained. In addition, the means of S_m are estimated, while the means of I_m are fixed to zero along the lines of longitudinal growth CFA models, see Mplus User's Guide, Muthén and Muthén (1998–2020), example 6.14. The loading parameters are unconstrained as in EFA, but are invariant across time to ensure that the factor measurements are consistent across time. The intercepts ν_p of the indicators are also time invariant. The residual variances of ε_{tp} are estimated as time invariant, although this restriction is generally not needed. The factor residuals ξ_{tm} are correlated within each time point but uncorrelated across time points. In addition, the factor variances of ξ_{tm} are fixed to 1 for $t = 0$ but are unconstrained at all other time points. This constraint is along the lines of longitudinal EFA/CFA analysis with invariant measurement structure. Model (11–12) is an example of EFA rotation that is neither oblique or orthogonal. The factor variance covariance is structured.

Figure 4 contains the Mplus input file for a simulation study for this model using $T = 5$, $P = 10$ and $M = 2$. For brevity we omitted the data generating model which is identical to the estimated model. Figure 5 contains the results of the simulation study for a selection of the parameters. The bias is minimal and the coverage is near the nominal level. Figure 6 shows the simplified model statement that can be used with real data. The third argument in the Geomin penalty function refers to the small ϵ used in the definition of the Geomin rotation function, for details see Asparouhov and Muthén (2024). Smaller values yield less bias but also yield a more erratic function to optimize. Typically, ϵ is set to 0.001 (default) or 0.0001.

6. Latent Variables as EFA Model Indicators

In a typical EFA model a set of observed variables measures a set of latent factors. In some circumstances however it might be of interest to use latent variables as indicators in an EFA model. These latent indicator variables will have their own separate measurement model, which may be an EFA or a CFA model. Such models have been discussed in Asparouhov and Muthén (2024) in the context of hierarchical EFA models where all the indicators are latent: EFA measured factors are used as measurement indicators for secondary factors. Here we focus on CFA measured latent variables that are used as indicators in an EFA model. In addition to the latent variable indicators, the EFA model also contains observed indicators.

In the ESEM framework implemented in Mplus, latent variables cannot be used as EFA indicators. However, latent variables can be regressed on EFA factors and thus can be viewed as indicators to the EFA factors, although these will not contribute to the EFA factor rotation. To be more specific, suppose that η_1 is a vector of latent variables measured by a vector of observed variables Z and η_2 is a vector of latent variables measured by a vector of observed variables Y and η_1. In this model, η_1 is measured by a CFA model while η_2 is measured by an EFA model. The model is given as follows

$$Z = \nu_1 + \Lambda_1\eta_1 + \varepsilon_1 \quad (13)$$

$$Y = \nu_2 + \Lambda_2\eta_2 + \varepsilon_2 \quad (14)$$

$$\eta_1 = \Lambda_3\eta_2 + \varepsilon_3 \quad (15)$$

```
                              ESTIMATES              S. E.     M. S. E.  95%  % Sig
                 Population   Average   Std. Dev.   Average              Cover Coeff

 F1       BY
  Y1                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  Y2                  0.800     0.7914     0.0434     0.0446     0.0019 0.920 1.000
  Y3                  0.400     0.3988     0.0281     0.0313     0.0008 0.970 1.000
  Y4                  0.700     0.6969     0.0417     0.0399     0.0017 0.910 1.000

 F1       ON
  F2                  0.500     0.5053     0.0329     0.0358     0.0011 0.990 1.000

 F2       ON
  X1                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  X2                  0.200     0.2084     0.0774     0.0761     0.0060 0.960 0.800
  X3                  1.100     1.1140     0.0977     0.1133     0.0096 0.990 1.000
  X4                  0.300     0.3014     0.0780     0.0782     0.0060 0.970 0.970
  X5                  0.400     0.3926     0.0853     0.0803     0.0073 0.950 1.000

 Y1       ON
  F2                  0.000    -0.0082     0.0137     0.0133     0.0003 1.000 0.000

 Y2       ON
  F2                  0.000     0.0050     0.0116     0.0140     0.0002 1.000 0.000

 Y3       ON
  F2                  0.200     0.1831     0.0236     0.0295     0.0008 0.950 1.000

 Y2       ON
  X1                  0.000    -0.0068     0.0189     0.0194     0.0004 1.000 0.000
  X2                  0.000    -0.0004     0.0340     0.0314     0.0011 0.950 0.050
  X3                  0.400     0.3887     0.0374     0.0427     0.0015 0.990 1.000
  X4                  0.000     0.0000     0.0338     0.0318     0.0011 0.960 0.040
  X5                  0.000    -0.0013     0.0318     0.0311     0.0010 0.950 0.050


 Residual Variances
  F1                  1.000     1.0112     0.0676     0.0750     0.0046 0.950 1.000
  F2                  0.000     0.0000     0.0000     0.0000     0.0000 1.000 0.000
```

Figure 2. Formative factor simulation study results.

```
MODEL:
        f1 BY y1-y4;
        f2 ON x1@1 x2-x5; f2 by; f2@0;
        f1 on f2;
        y1-y4 on f2 (a1-a4);
        y1-y3 on x1-x5 (b1-b15);

MODEL PRIORS:
        a1-a4~ALF(0,1);
        b1-b15~ALF(0,1);
```

Figure 3. Formative factor analysis with PSEM.

where Λ_1 is a structured CFA loading structure, while Λ_2 and Λ_3 are unconstrained EFA loading structures. In the ESEM framework the above model can be estimated by selecting an optimal rotation for η_2 which simplifies Λ_2. Here we will show how PSEM can be used to estimate the model by selecting an optimal rotation for η_2 which simplifies both Λ_2 and Λ_3.

Before we proceed, however, we want to discuss the justification for including Λ_3 in the rotation selection. If the latent variables η_1 are expected to perform as "pure" indicators to a large extent, i.e., similarly to the observed indicators Y, then it makes sense that the rotation should be chosen to simplify not just Λ_2 but also Λ_3. If Λ_3 is not expected to be close to a simple loading structure but is likely to contain only non-zero values, then including that matrix in the rotation selection would be counterproductive and may result in less accurate results. Thus, context and substantive judgment are important in choosing between the ESEM and PSEM models. If the size of Λ_3 however is substantially smaller than the size of Λ_2, we can expect that the methods will be quite similar as Λ_2 is likely to dominate the rotation selection optimization.

To illustrate the PSEM estimation we conduct a simulation study using 2 latent indicators and 2 EFA factors, i.e., both η_1 and η_2 are of size 2. The indicator vectors Y and Z are of size 6. In this simulation, Λ_2 has two cross-loadings while Λ_3 has none. Figure 7 contains the simulation study using the PSEM method and Figure 8 contains the simulation study using the ESEM method. Note that η_1 latent variables are uncorrelated. Since these are indicators for η_2, their correlation is modeled by the EFA model for η_2. The results of the simulation study for the PSEM method are given in Figure 9. The bias in the parameter estimates is small and the coverage is near the nominal levels. The ESEM results for this simulation are similar but are worse than those of the PSEM method, in terms of bias and MSE.

```
montecarlo:
        names = y1-y50;
        nobs = 1000;
        nreps = 100;

model:
    f01 by y1-y5*1 y6*0.5 y7-y10*0 (a1-a10);
    f02 by y1-y5*0 y6-y10*1 (a11-a20);
    f11 by y11-y15*1 y16*0.5 y17-y20*0 (a1-a10);
    f12 by y11-y15*0 y16-y20*1 (a11-a20);
    f21 by y21-y25*1 y26*0.5 y27-y30*0 (a1-a10);
    f22 by y21-y25*0 y26-y30*1 (a11-a20);
    f31 by y31-y35*1 y36*0.5 y37-y40*0 (a1-a10);
    f32 by y31-y35*0 y36-y40*1 (a11-a20);
    f41 by y41-y45*1 y46*0.5 y47-y50*0 (a1-a10);
    f42 by y41-y45*0 y46-y50*1 (a11-a20);
    f01-f02@1 f11*0.9 f12*1.1 f21*0.9 f22*1;
    f31*0.8 f32*0.8 f41*1.1 f42*1.2;
    f01 with f02*0.2;
    f11 with f12*0.2;
    f21 with f22*0.3;
    f31 with f32*0.3;
    f41 with f42*0.3;
    i1 s1 | f01@0 f11@1 f21@2 f31@3 f41@4;
    i2 s2 | f02@0 f12@1 f22@2 f32@3 f42@4;
    [i1@0 s1*1 i2@0 s2*-0.3];
    i1*1; s1*.15; i1 with s1*.1;
    i2*0.8; s2*.1; i2 with s2*.05;
    i1 with i2*0.3; s1 with s2*0.05;
    [y1-y10*0] (m1-m10);
    [y11-y20*0] (m1-m10);
    [y21-y30*0] (m1-m10);
    [y31-y40*0] (m1-m10);
    [y41-y50*0] (m1-m10);
    y1-y10*1 (v1-v10);
    y11-y20*1 (v1-v10);
    y21-y30*1 (v1-v10);
    y31-y40*1 (v1-v10);
    y41-y50*1 (v1-v10);

model prior: a1-a20~Geomin(2,1,0.0001);
```

Figure 4. EFA growth simulation study.

This is explained by the fact that the additional pure latent indicators contribute to the optimal rotation selection. On the other hand, if we generate data where Λ_3 contains only non-zero values, i.e., 2 cross-loadings, then the situation is reversed and the ESEM method performs slightly better than the PSEM method. This is again explained with the fact that the added latent indicators contribute more cross-loadings and thus hinder the recovery of the original structure. It should be noted here that the meaning of bias is only in the context of recovering simple structure and not in terms of standard model estimation bias. As with every EFA estimation, the more cross-loadings there are, the less likely the generating parameters are to be recovered in its original form, and the more likely it is that a simpler structure, equally well fitting, will be found as a replacement.

The question of which model parts should contribute to the optimal factor rotation arises in other contexts as well. Consider the ESEM model where EFA measured factors η are regressed on covariates X

$$Y = \nu + \Lambda\eta + \varepsilon \tag{16}$$

$$\eta = BX + \xi. \tag{17}$$

The question we consider is this: should the optimal factor rotation be selected to simplify not just Λ but also B. From a practical perspective it will be somewhat harder to justify a simple form for B than it is for Λ but let's ignore that point for now. Suppose that B has indeed a simple form and each of the covariates predicts just one factor. Would including the B matrix in the rotation function strengthen the optimal rotation selection. It turns out that the answer is no. The two matrices are rotated in an opposite direction which is why the information does not combine easily. Even when B has a simple structure, based on simulation studies that we do not report, it is the case that selecting the optimal rotation based on simplifying Λ alone, B alone, or Λ and B together works about equally well. Thus, utilizing PSEM here for the purpose of simplifying more than just the measurement loading matrix is not beneficial. ESEM, which only works on simplifying Λ works equally well. Furthermore, if we consider the case of non-simple B structures, simulations not reported here show that there is a substantial drawback if the optimal rotation attempts to simplify both Λ and B, and PSEM results can be substantially worse than ESEM. Unlike the case of latent EFA indicators, expanding the model simplification goals beyond the loading matrix in a model like (16–17) is not recommended at this time.

7. Measurement Invariance in Latent Transition Analysis

The PSEM modeling framework is defined generally to make an unidentified model into an identified model via the addition of a penalty function which is optimized simultaneously with the likelihood. The log-likelihood of the PSEM model is identical to the null model, i.e., an identified version of the model where a set of unidentified parameters are fixed. The PSEM model is generally an interpretable model that is practically desirable while the null model is generally an impractical model that lacks proper interpretation. For example, in the EFA settings, the PSEM model is the rotated EFA model while the null model is the unrotated CFA model where the loadings above the diagonal are fixed to zero. Similarly in multiple group factor analysis, the PSEM model is the aligned model which allows factor means comparison across groups while the null model is the configural model where factor means are all fixed to zero and group comparison is not available. Prior to the introduction of the PSEM framework, the Regularized SEM (RegSEM) framework was introduced in Jacobucci et al. (2016) following the idea of regularized LASSO regression, Tibshirani (1996). The difference between PSEM and regularized models is that in RegSEM, the penalty function may alter the log-likelihood, i.e., a small portion of the likelihood is sacrificed to obtain a smaller penalty and a more interpretable model. In proper PSEM applications, the log-likelihood is not sacrificed for the benefit of minimizing the penalty. The penalty is minimized only within the space of unidentified dimensions. In terms of algorithms and software implementation,

			ESTIMATES		S. E.	M. S. E.	95%	% Sig
		Population	Average	Std. Dev.	Average		Cover	Coeff
F01	BY							
Y1		1.000	1.0032	0.0380	0.0398	0.0014	0.940	1.000
Y2		1.000	1.0025	0.0381	0.0397	0.0014	0.940	1.000
Y3		1.000	1.0037	0.0380	0.0398	0.0014	0.920	1.000
Y4		1.000	1.0024	0.0367	0.0398	0.0013	0.940	1.000
Y5		1.000	1.0032	0.0369	0.0398	0.0014	0.950	1.000
Y6		0.500	0.4980	0.0192	0.0209	0.0004	0.970	1.000
Y7		0.000	-0.0026	0.0052	0.0055	0.0000	0.970	0.030
Y8		0.000	-0.0032	0.0046	0.0055	0.0000	0.970	0.030
Y9		0.000	-0.0028	0.0046	0.0055	0.0000	0.990	0.010
Y10		0.000	-0.0026	0.0047	0.0055	0.0000	0.980	0.020
F02	BY							
Y1		0.000	-0.0007	0.0076	0.0079	0.0001	0.980	0.020
Y2		0.000	0.0006	0.0091	0.0081	0.0001	0.940	0.060
Y3		0.000	-0.0001	0.0076	0.0079	0.0001	0.980	0.020
Y4		0.000	-0.0010	0.0087	0.0082	0.0001	0.980	0.020
Y5		0.000	-0.0008	0.0080	0.0079	0.0001	0.970	0.030
Y6		1.000	1.0066	0.0415	0.0412	0.0017	0.940	1.000
Y7		1.000	1.0065	0.0429	0.0412	0.0019	0.950	1.000
Y8		1.000	1.0077	0.0438	0.0412	0.0020	0.920	1.000
Y9		1.000	1.0068	0.0428	0.0412	0.0019	0.940	1.000
Y10		1.000	1.0070	0.0425	0.0412	0.0018	0.920	1.000
F01	WITH							
F02		0.200	0.2076	0.0539	0.0537	0.0029	0.940	0.950
Means								
I1		0.000	0.0000	0.0000	0.0000	0.0000	1.000	0.000
S1		1.000	0.9980	0.0421	0.0425	0.0018	0.950	1.000
I2		0.000	0.0000	0.0000	0.0000	0.0000	1.000	0.000
S2		-0.300	-0.2942	0.0206	0.0192	0.0005	0.930	1.000
Variances								
I1		1.000	1.0000	0.1443	0.1311	0.0206	0.930	1.000
S1		0.150	0.1483	0.0179	0.0198	0.0003	0.960	1.000
I2		0.800	0.7905	0.1204	0.1182	0.0145	0.950	1.000
S2		0.100	0.0998	0.0158	0.0160	0.0002	0.930	1.000
Residual Variances								
F11		0.900	0.8985	0.0900	0.0920	0.0080	0.970	1.000
F12		1.100	1.0857	0.1174	0.1085	0.0139	0.920	1.000
F21		0.900	0.8959	0.0826	0.0938	0.0068	0.970	1.000
F22		1.000	0.9781	0.0978	0.1026	0.0099	0.950	1.000
F31		0.800	0.8036	0.0950	0.0914	0.0089	0.940	1.000
F32		0.800	0.7927	0.0976	0.0913	0.0095	0.910	1.000

Figure 5. EFA growth simulation study results.

however, there is no difference between the two methods. In Mplus, simply using a penalty with an identified model leads to a RegSEM model.

The Mplus framework is built along the lines that continuous and categorical latent variables are used to model relations between observed variables. Models with continuous latent variables (SEM models) have analogue models that use categorical latent variables (Mixture models). However, SEM models typically fit only the first and second order moments of the observed data, while finite mixture models fit also higher level moments. Because of that, a continuous latent variable model that is unidentified, might have a categorical latent variable analogue that is identified. Thus, unidentified models that PSEM takes advantage of are not as easily available in the Mixture settings.

As an example consider an LTA model with time invariant latent class measurement as well as the "continuous" variable equivalent: longitudinal factor analysis with invariant measurement. In the continuous case, the measurement invariance in the model allows us to estimate time specific factor mean and if the measurement invariance is relaxed the model becomes unidentified. This makes it suitable for PSEM. In the LTA case, the time-invariant measurement model allows us to estimate time specific distribution for the latent class variable, however, relaxing the measurement invariance does not result in an unidentified model. Therefore adding a penalty function which forces invariance as much as possible will result in a RegSEM model rather than a PSEM model. Nevertheless, here we will explore the advantages of these RegSEM mixture models that are

```
model:
 f01-f02 by y1*1 y2-y10 (a1-a20);
 f11-f12 by y11*1 y12-y20 (a1-a20);
 f21-f22 by y21*1 y22-y30 (a1-a20);
 f31-f32 by y31*1 y32-y40 (a1-a20);
 f41-f42 by y41*1 y42-y50 (a1-a20);
 i1 s1 | f01@0 f11@1 f21@2 f31@3 f41@4;
 i2 s2 | f02@0 f12@1 f22@2 f32@3 f42@4;
 f01-f02@1;
 f01 with f02;
 f11 with f12;
 f21 with f22;
 f31 with f32;
 f41 with f42;
 [y1-y10] (m1-m10);
 [y11-y20] (m1-m10);
 [y21-y30] (m1-m10);
 [y31-y40] (m1-m10);
 [y41-y50] (m1-m10);
 y1-y10 (v1-v10);
 y11-y20 (v1-v10);
 y21-y30 (v1-v10);
 y31-y40 (v1-v10);
 y41-y50 (v1-v10);

model prior: a1-a20~Geomin(2,1,0.0001);
```

Figure 6. EFA growth modeling with PSEM.

```
montecarlo:
names = y1-y6 z1-z6;
nobs = 500;
nreps = 100;

model montecarlo:
f1 by z1@1 z2-z3*1;
f2 by z4@1 z5-z6*.8;
z1-z6*1; f1-f2*1; f1 with f2@0;

f3 by y1-y3*1 y4*0.2 y5-y6*0 f1*1 f2*0;
f4 by y1-y2*0 y3*0.3 y4-y6*1 f1*0 f2*0.9;
y1-y6*1; f3-f4@1;  f3 with f4*0.4;


model:
f1 by z1@1 z2-z3*1;
f2 by z4@1 z5-z6*.8;
z1-z6*1; f1-f2*1; f1 with f2@0;

f3 by y1-y3*1 y4*0.2 y5-y6*0 f1*1 f2*0 (L1-L8);
f4 by y1-y2*0 y3*0.3 y4-y6*1 f1*0 f2*0.9 (L9-L16);
y1-y6*1; f3-f4@1;  f3 with f4*0.4;

model prior:
L1-L16~geomin(2,1,0.0001);
```

Figure 7. EFA with latent indicators: PSEM method.

constructed along the lines of their continuous variable counterparts. Regularized Mixture models have also been discussed in Shedden and Zucker (2008).

To illustrate this concept we consider a latent class variable measured at 3 time points by 4 binary variables as in an LTA model. Measurement invariance holds only partially as it is often the case in real data. We generate the data so that measurement invariance holds for 7 out of the 8 measurement parameters at each time point, i.e., there are 3 non-invariant measurement parameters. We consider the following 3 models. The first model is the LTA-PSEM model where a measurement invariance penalty is added to the model. As we discussed earlier this is in fact a RegSEM model. The second model is the model where measurement invariance is removed. The third model is the LTA model which forces measurement invariance incorrectly. Figure 10 contains the Mplus input file for this simulation study for the LTA-PSEM model. The model population part of the input is omitted for brevity but is identical to the model statement. The parameter estimates for the PSEM-LTA are unbiased and the coverage is near the nominal levels. In that regard, the PSEM-LTA model can be used to determine which measurement parameters are invariant and which are not. Figure 11 contains the results for all three models for a selection of the parameters, namely the time specific class distribution. Here we see that PSEM-LTA outperforms the other two estimations substantially in terms of MSE. In addition, assuming measurement invariance leads to bias in the parameter estimates and lower coverage. The addition of the measurement invariance penalty leads to substantial benefits for the estimation without damaging the data fit. The drop in the log-likelihood due to the penalty addition is 1 on average across the 100 replications. As a comparison, the drop in the likelihood resulting from strictly imposing measurement invariance is close to 100 on average. This also leads to LRT rejection for the measurement invariance. The PSEM-LTA model is essentially estimating the LTA model with approximate measurement invariance.

```
montecarlo:
names = y1-y6 z1-z6;
nobs = 500;
nreps = 100;

model montecarlo:
f1 by z1@1 z2-z3*1;
f2 by z4@1 z5-z6*.8;
z1-z6*1; f1-f2*1; f1 with f2@0;

f3 by y1-y3*1 y4*0.2 y5-y6*0 f1*1 f2*0;
f4 by y1-y2*0 y3*0.3 y4-y6*1 f1*0 f2*0.9;
y1-y6*1; f3-f4@1;  f3 with f4*0.4;


model:
f1 by z1@1 z2-z3*1;
f2 by z4@1 z5-z6*.8;
z1-z6*1; f1-f2*1; f1 with f2@0;

f3 by y1-y3*1 y4*0.2 y5-y6*0  (*1);
f4 by y1-y2*0 y3*0.3 y4-y6*1  (*1);
y1-y6*1; f3-f4@1;  f3 with f4*0.4;

f1 on f3*1 f4*0;
f2 on f3*0 f4*0.9;
```

Figure 8. EFA with latent indicators: ESEM method.

In this simulation study the measurement non-invariance is small and easily detectable. In practical applications that might not be the case. The weight of the penalty will likely need further analysis in such situations. Multiple models may need to be considered with different levels of penalty

		ESTIMATES		S. E.	M. S. E.	95%	% Sig
	Population	Average	Std. Dev.	Average		Cover	Coeff
F1 BY							
Z1	1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
Z2	1.000	1.0006	0.0592	0.0529	0.0035	0.940	1.000
Z3	1.000	1.0006	0.0569	0.0531	0.0032	0.910	1.000
F2 BY							
Z4	1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
Z5	0.800	0.8070	0.0573	0.0547	0.0033	0.940	1.000
Z6	0.800	0.7966	0.0566	0.0541	0.0032	0.910	1.000
F3 BY							
Y1	1.000	1.0018	0.0690	0.0691	0.0047	0.950	1.000
Y2	1.000	1.0058	0.0671	0.0708	0.0045	0.970	1.000
Y3	1.000	1.0027	0.0730	0.0773	0.0053	0.960	1.000
Y4	0.200	0.1726	0.0835	0.0816	0.0077	0.900	0.510
Y5	0.000	-0.0138	0.0453	0.0501	0.0022	1.000	0.000
Y6	0.000	-0.0187	0.0493	0.0542	0.0028	0.980	0.020
F4 BY							
Y1	0.000	-0.0074	0.0349	0.0494	0.0013	1.000	0.000
Y2	0.000	-0.0012	0.0492	0.0559	0.0024	1.000	0.000
Y3	0.300	0.2935	0.0741	0.0827	0.0055	0.960	0.930
Y4	1.000	1.0048	0.0815	0.0779	0.0066	0.930	1.000
Y5	1.000	0.9974	0.0688	0.0700	0.0047	0.970	1.000
Y6	1.000	0.9908	0.0704	0.0709	0.0050	0.980	1.000
F3 BY							
F1	1.000	1.0023	0.0839	0.0827	0.0070	0.970	1.000
F2	0.000	-0.0102	0.0529	0.0615	0.0029	0.990	0.010
F4 BY							
F1	0.000	-0.0088	0.0483	0.0602	0.0024	0.990	0.010
F2	0.900	0.8970	0.0786	0.0820	0.0061	0.950	1.000
F3 WITH							
F4	0.400	0.4097	0.0591	0.0675	0.0036	0.990	1.000

Figure 9. EFA with latent indicators: PSEM results.

weight (grid search). The different PSEM-LTA models may need to be converted to non-PSEM models where approximately invariant parameters are actually held equal and the non-invariant parameters are estimated as non-invariant. The BIC criterion for the non-PSEM models can then be used for the final model selection.

8. Direct Effects in Latent Class Analysis

In this section we provide another illustration where a continuous variable PSEM model suggests applications for RegSEM Mixture modeling. First consider the continuous variable MIMIC model where a latent variable is predicted by a covariate but the covariate may also have a direct effect on some of the indicators. The PSEM model allows us to include all direct effects with LASSO/ALF priors which forces the direct effects to stay near zero unless the data fit mandates a non-zero value. In this process, the predictive power of the covariate on the latent variable remains intact. Without the prior/penalty, including all direct effects and the effect of the covariate on the factor is an unidentified model. The PSEM framework takes advantage of that. In LCA, the equivalent situation where all direct effects are included in addition to the effect of the covariate on the latent class variable is a generally identified model. This suggests that the MIMIC-equivalent approach, i.e., adding LASSO/ALF priors for the direct effects in LCA, will be a RegSEM Mixture model where a portion of the likelihood is sacrificed to the benefit of the penalty. In all such applications, it is important to make sure that the penalty is not weighted too heavily so that the overall optimization does not alter the log-likelihood substantially. As in the previous example, here we expect that the added penalty will benefit the Mixture model by stabilizing the estimation and would reduce MSE of the estimates.

We illustrate this situation using a simulation study where 6 binary indicators measure a 2-class latent variable. A latent class predictor is included in the model which also has 3 direct effects to the indicators, i.e., 3 direct effects are non-zero and 3 direct effects are zero. Direct effects in Mixture models can be class specific which corresponds to a latent class and covariate interaction. Such effects are

```
montecarlo:
 names are u11-u14 u21-u24 u31-u34;
 genclasses = c1(2) c2(2) c3(2);
 classes = c1(2) c2(2) c3(2);
 generate = u11-u34(1);
 categorical = u11-u34;
 nobs = 1000;
 nrep = 100;

analysis: type = mixture;

model:
 %overall%
 [c1#1*0.4 c2#1*-0.7 c3#1*0.8];
 C2#1 on C1#1*0.5;
 C3#1 on C2#1*-0.3;

model c1:
 %c1#1%
 [u11$1-u14$1*-1.3] (a1-a4);
 %c1#2%
 [u11$1-u13$1*1.3 u14$1*0.3] (a5-a8);

model c2:
 %c2#1%
 [u21$1-u24$1*-1.3] (b1-b4);
 %c2#2%
 [u21$1*0.5 u22$1-u24$1*1.3] (b5-b8);

model c3:
 %c3#1%
 [u31$1-u32$1*-1.3 u33$1*0 u34$1*-1.3] (c1-c4);
 %c3#2%
 [u31$1-u34$1*1.3] (c5-c8);

model prior:
do(#,1,8) DIFF(a# b# c#)~ALF(0,1);
```

Figure 10. PSEM-LTA simulation study.

secondary in nature and are less common in practice but certainly can occur. For this simulation study, we consider only class invariant direct effects. The conclusions of the simulation study, however, carry over also to the class specific direct effects.

In this model estimation, all direct effects are included. This corresponds to the situation where we do not know which direct effects must be included in the model. All direct effects are given ALF priors to keep them near zero when possible. For comparison, we also include a simulation study where the penalty function is not included. Without the penalty, the model is identified, although in some cases the identification is weak. In this simulation study, we deliberately choose the smaller sample size of $N = 200$. This is because smaller sample sizes are more likely to need the stabilization power of the penalty. With an infinitely large sample size, the model without penalty is expected to perform well.

Mixture models are prone to multiple local solutions. This is even more so when the model is extra flexible, as is the case of including all direct effects. Smaller sample sizes also make multiple local solutions more likely. This is why we included in this simulation study random starting values. In real data analysis, random starting values are used by default in Mplus but in simulation studies they are not. When the classes in a model are equivalent in nature and each class has the same number of parameters, the order of the estimated classes is somewhat random. Random starting values may randomly reorder the classes. This is not of importance in real data analysis but in simulation studies it becomes an obstacle as we want to make sure that the average estimates across the simulations are computed using the same classes and the class ordering is the same across replications. Because we include random starting values in this simulation, we must also include a feature that would

```
                          ESTIMATES              S. E.     M. S. E.   95%  % Sig
              Population   Average   Std. Dev.   Average              Cover Coeff


PSEM-LTA

  C1#1            0.400     0.4004     0.1011     0.1276     0.0101 0.980 0.900
  C2#1           -0.700    -0.6933     0.1851     0.1859     0.0340 0.950 0.960
  C3#1            0.800     0.8033     0.1547     0.1774     0.0237 0.960 0.990


LTA without measurement invariance

  C1#1            0.400     0.3938     0.1325     0.1409     0.0174 0.960 0.830
  C2#1           -0.700    -0.6913     0.2038     0.1965     0.0412 0.960 0.940
  C3#1            0.800     0.8218     0.1923     0.1888     0.0371 0.940 0.970


LTA with measurement invariance

  C1#1            0.400     0.5597     0.0997     0.1122     0.0353 0.720 1.000
  C2#1           -0.700    -0.6519     0.1945     0.1806     0.0398 0.910 0.950
  C3#1            0.800     0.5914     0.1625     0.1550     0.0696 0.690 0.980
```

Figure 11. LTA simulation study results.

```
montecarlo:
 names are u1-u6 x;
 generate = u1-u6(1);
 categorical = u1-u6;
 nobs =200;
 nrep = 100;
 genclass=c(2);
 class=c(2);

analysis: type=mixture; starts=40 20;

model population:
    %overall%
    x*1;
    u1-u3 on x*0;
    u4 on x*1;
    u5 on x*1;
    u6 on x*-0.4;
    c on x*.5;
    [c#1*1];

    %c#1%
    [u1$1-u3$1*1 u4$1*1.2 u5$1*0 u6$1*-1.5];

    %c#2%
    [u1$1*-1.3 u2$1-u3$1*-1 u4$1*-0.4 u5$1-u6$1*1];

model:
    %overall%
    c on x*.5;
    u1-u3 on x*0 (a1-a3);
    u4 on x*1 (a4);
    u5 on x*1 (a5);
    u6 on x*-0.4 (a6);
    [c#1*1] (a);

    %c#1%
    [u1$1-u3$1*1 u4$1*1.2 u5$1*0 u6$1*-1.5];

    %c#2%
    [u1$1*-1.3 u2$1-u3$1*-1 u4$1*-0.4 u5$1-u6$1*1];

model prior: a1-a6~ALF(0,1);

model constraint: a>0;
```

Figure 12. PSEM-LCA with direct effects.

enforce the same class ordering across the replications. In this simulation study, we use a model constraint which automatically orders the larger class as the first class.

Figure 12 shows the Mplus input file for the PSEM-LCA simulation study. The regular LCA analysis uses the same setup except for the penalty specification given in MODEL PRIOR. The results for both model estimations are given in Figure 13. We include only the predictor effect estimates in this figure. The multiple **** here indicates that the value is too large to print in the allotted space. The stabilization effect of the penalty is clearly visible in this comparison. In addition, the MSE reduction is obtained in all regression parameters as expected. The LCA-PSEM model yields unbiased estimates and coverage near the nominal levels. The log-likelihood drop caused by the penalty inclusion is minimal. The average drop across the replications is 0.2. We conclude that the penalty inclusion can be beneficial when direct effects need to be explored. To a large extent, the PSEM-LCA method with all direct effects included can also be viewed as a method for discovering direct effects. It can also be viewed as an alternative to multistage estimation used in LCA, see Asparouhov and Muthén (2014a). The multistage estimation is primarily designed to prevent unwanted direct effects to alter the class formation. If the direct effects are properly accounted for in the latent class estimation, then the effect of the covariates on the latent class variable is unbiased.

9. Class Enumeration

The most fundamental question in finite mixture modeling is to determine the number of classes. The LRT test comparing the K and the $K+1$ class models does not have an explicit distribution and thus is impractical to use for this purpose. The BIC criterion is the most well performing tool that is easily available, see Nylund et al. (2007). In this section we explore the possibility to use PSEM for class enumeration. The idea is as follows. We introduce a penalty for each additional class that is needed to fit the data well. This way small classes will be eliminated to avoid the penalty. Classes that can be combined will be combined to avoid the penalty. Adding a penalty for every additional class that is needed is similar to the Chinese restaurant process used in Bayesian analysis, see Gelman et al. (2004). The penalty essentially acts as a prior and thus the similarities are clear. It is possible to choose a penalty that has a progressively higher penalty for every additional class. We will not pursue this here and will instead have a constant penalty for every additional class. The precise value of the penalty for each class will be computed to match the BIC criterion. Since minimizing BIC leads to the correct number of classes asymptotically, the penalty should match the BIC penalty for each additional class. The BIC criterion is given by

$$BIC = P \cdot \log(N) - 2 \cdot LL \tag{18}$$

where P is the number of model parameters, N is the sample size, and LL is the log-likelihood. This means that the BIC criterion penalty for each additional parameter is $\log(N)/2$. Suppose that each additional class contributes P_0 parameters in the model. Then the penalty for each additional class should be $P_0 \log(N)/2$. That is, in order for us to allow an additional class, the log-likelihood of the model must be improved by at least $P_0 \log(N)/2$. There are a number of ways to implement such a penalty. For simplicity, we shall use a penalty prior for the parameters in the distribution of the latent class variable. Suppose that the latent class variable distribution is given by

$$P(C=k) = \frac{Exp(\alpha_k)}{\sum_k Exp(\alpha_k)} \tag{19}$$

for $k = 1, \ldots, K$, where K is the total number of classes. Here α_K is fixed to 0 for identification purposes. If α_k is a large negative number such as -15, the class will be empty because $P(C=k)$ will be zero. We introduce a penalty for the parameters α_k as $ALF(-15, \nu)$. The penalty is thus

```
                                ESTIMATES              S. E.     M. S. E.  95%  % Sig
                 Population   Average   Std. Dev.   Average             Cover Coeff

 PSEM-LCA

  U1          ON
   X                 0.000    -0.0120     0.1917     0.1907     0.0365 0.960 0.040
  U2          ON
   X                 0.000    -0.0365     0.2181     0.1925     0.0484 0.950 0.050
  U3          ON
   X                 0.000    -0.0019     0.2202     0.2050     0.0480 0.960 0.040
  U4          ON
   X                 1.000     1.0251     0.2467     0.2582     0.0609 0.940 0.980
  U5          ON
   X                 1.000     1.0062     0.2037     0.2119     0.0411 0.940 1.000
  U6          ON
   X                -0.400    -0.3912     0.2809     0.2685     0.0782 0.840 0.270
  C#1         ON
   X                 0.500     0.4882     0.3743     0.3434     0.1388 0.940 0.360


 Regular LCA

  U1          ON
   X                 0.000    -0.0331     0.2740     0.2336     0.0754 0.900 0.100
  U2          ON
   X                 0.000    -0.0762     0.2999     0.2226     0.0948 0.880 0.120
  U3          ON
   X                 0.000    -0.0389     0.2879     0.2293     0.0836 0.880 0.120
  U4          ON
   X                 1.000     1.0318     0.2658     0.2489     0.0709 0.910 0.970
  U5          ON
   X                 1.000     1.0379     0.2166     0.2144     0.0479 0.920 1.000
  U6          ON
   X                -0.400   -96.3775   960.0707     0.2785 ********** 0.860 0.410
  C#1         ON
   X                 0.500     0.4306     0.4892     0.3435     0.2417 0.850 0.450
```

Figure 13. LCA with direct effects results.

$\sqrt{\alpha_k + 15}/v$. We need to determine the value of v to match the BIC penalty.

The last class doesn't have a penalty. It will be the default for the observations and is expected to be the largest class. Here we will make an assumption that a class which is smaller than 1% of the largest class would be considered too small for practical purposes. A class which is 1% of the largest class has a α_k parameter of -4.6 which is $\log(0.01)$. We determine v by setting the BIC penalty to be equal to the ALF penalty at $\alpha_k = -4.6$. Classes that are larger than the 1% of the largest class will have a little bigger penalty but these classes will also be bigger in size so presumably the sample size will compensate for that discrepancy. Because the square root is not a flat function, the penalty increases slightly for the larger classes but not by more than $20\% = \sqrt{(15-0)/(15-4.6)} - 1$. This is an advantage of the ALF prior (square root) as it resembles a flat line in the interval (10.4,15) more so than LASSO or Normal prior.

We now obtain the following equation for v by setting the ALF penalty be equal to the BIC penalty

$$\sqrt{10.4}/v = P_0 \log(N)/2 \tag{20}$$

or

$$v = \frac{6.45}{P_0 \log(N)}. \tag{21}$$

Classes that are smaller than the 1% of the largest class will be considered empty. In most cases, however, because we are utilizing alignment priors, we expect that empty classes will all have $\alpha_k = -15$.

We select K as the maximum number of classes that can occur for the model. For example, if we are not expecting more than 10 classes to be estimated by the finite mixture PSEM model we can estimate a PSEM model with $K = 10$ classes and determine the number of classes that are not empty. The class enumeration technique will then conclude that the proper number of classes for this model is the number of non-empty classes found in this PSEM Mixture model.

Next, we illustrate this methodology with a simulation study using a 4-class latent profile analysis with 10 indicators and $N = 500$. We use a 10 class estimation, i.e., $K = 10$ and we simultaneously explore finite mixture models with up to 10 classes. In this model, each class contributes an additional 11 parameters and so $P_0 = 11$ which leads to $v = 0.094$. Figure 14 shows the input file for this simulation study. Note that the large amount of random starting values is key for this analysis. Figure 15 shows the results of the analysis. There are 6 empty classes where $\alpha_k = -15$ and thus the analysis correctly concludes that the proper number of classes for this model is 4. For comparison, if the 10-class model is estimated without the prior, all 10 classes are above the 1% threshold. Using the BIC criterion directly also concludes that there are 4 classes. Repeating this analysis 10 times yields the same result in each replication. It should be noted that because the non-empty classes do not appear in the same position across the replications, it is not as easy to conduct enumeration simulation studies without some additional techniques. It is possible to order the classes by size but that will hinder the random starting values to some extent. With the current setup, the results of each replication must be saved and manually evaluated.

```
montecarlo:
 names are y1-y10;
 genclasses = c(4);
 classes = c(10);
 nobs = 500;
 nrep = 1;

analysis:
 type = mixture;
 STARTS=2000 500;

model population:
 %overall%
 y1-y10*1;
 %c#1%
 [y1-y10*-1];
 %c#2%
 [y1-y10*0];
 %c#3%
 [y1-y10*1];
 %c#4%
 [y1-y10*2];

model:
 %overall%
 [c#1-c#9] (p1-p9);

model prior:
  p1-p9~ALF(-15,0.094);
```

Figure 14. PSEM class enumeration.

10. Finite Mixtures of Exploratory Structural Equation Models (ESEM)

New models can be obtained also by combining finite mixture modeling with the standard PSEM models described in Asparouhov and Muthén (2024). In this section we describe one such example: Mixture of ESEM models. It is already possible to estimate Mixtures of EFA models in Mplus directly without the use of PSEM. EFA models however do not have the full flexibility of ESEM models. In this illustration we add a covariate which predicts the factors in the EFA model. This model is a pure PSEM model. The null model is the finite mixture of unrotated ESEM models. The penalty function as usual is not expected to affect the data fit and the weight is determined numerically: the penalty weight is the largest value which leads to the same log-likelihood as the null model log-likelihood.

The Mixture-ESEM model extracts subpopulations in the data, where the observed variables within each subpopulation are modeled with an ESEM model. The ESEM model can have a class-invariant measurement structure, a class-specific measurement structure, or it can have different numbers of factors in each class. Here we consider continuous class indicators. Thus, this mixture model can be viewed as latent profile analysis (LPA) with structured residual covariance within class. In the absence of class-invariant EFA/CFA structure, the intercepts/means of the indicators are class specific. Furthermore, these intercepts are the driving force in the latent class formation. Class specific differences in the ESEM model can serve as additional information that can identify the latent subpopulations but typically the means of the observed variables is the main class differentiator as in LPA. Also, the latent class variable is typically the main explanatory variable for covariance between the observed variables. The factor analysis within class is typically secondary in terms of explaining the covariances between the variables. As such, the within class factor model is often unknown and can benefit from exploratory techniques and the flexibility of ESEM.

```
Categorical Latent Variables

 Means
  C#1             0.000   -15.0000    0.0000    0.0000   224.9991 0.000 1.000
  C#2             0.000    -0.1791    0.0000    0.1438     0.0321 1.000 0.000
  C#3             0.000    -0.2522    0.0000    0.1577     0.0636 1.000 0.000
  C#4             0.000    -0.1428    0.0000    0.1563     0.0204 1.000 0.000
  C#5             0.000   -15.0000    0.0000    0.0000   224.9997 0.000 1.000
  C#6             0.000   -15.0000    0.0000    0.0000   224.9999 0.000 1.000
  C#7             0.000   -15.0000    0.0000    0.0000   224.9999 0.000 1.000
  C#8             0.000   -14.9998    0.0000    0.0003   224.9931 0.000 1.000
  C#9             0.000   -14.9998    0.0000    0.0003   224.9931 0.000 1.000
```

Figure 15. PSEM class enumeration results.

```
montecarlo:
 names are y1-y8 x;
 genclasses = c(2);
 classes = c(2);
 nobs = 2000;
 nrep = 100;

analysis: type = mixture;

model:
 %overall%
 f1 by y1-y4*1 y5*0.5 y6-y8*0 (a1-a8);
 f2 by y1*0.2 y2-y4*0 y5-y8*.5  (a9-a16);
 [y1*1 y2*1 y3*1 y4*1 y5*0 y6*0 y7*0 y8*0];
 f1 with f2*0.4;
 f1 on x*0.4;
 f2 on x*-0.2;
 y1-y8*.5;
 f1-f2@1;
 [f1-f2@0];
 c#1 on x*0.5;
 [c#1*0];

 %c#2%
 f1 by y1-y4*.5 y5*0.3 y6-y8*0 (b1-b8);
 f2 by y1*0.2 y2-y4*0 y5-y8*1 (b9-b16);
 [y1*0 y2*0 y3*0 y4*0 y5*1 y6*1 y7*1 y8*1];
 f1 with f2*0.2;
 f1 on x*0.2;
 f2 on x*0.4;

model prior:
a1-a16~geomin(2,1,.0001);
b1-b16~geomin(2,1,.0001);
```

Figure 16. ESEM mixture simulation study.

Figure 16 contains the input file for a 2-class 2-factor ESEM model with 8 continuous endogenous variables and one covariate. For brevity the model population statement is omitted but is identical to the model statement. The means of the dependent variables are class specific. The EFA structure is also class specific. The results of this simulation study for a selection of the parameters is given in Figure 17. The bias in the parameter estimates is minimal and the coverage is near the nominal levels.

In this simulation study the entropy for the mixture model is 0.6. This level of entropy is generally considered to be in the moderate range. At this level, latent classes are not easily differentiated and class membership is typically based on the entirety of the observed data. There are no elements in the model and the data that can easily determine the class membership. In such situations, the within class model specification becomes important and class formation can be affected if the within class model changes. For this particular data, there are two other alternative models that do not use ESEM for the within class model. The first model is the LPA model with unrestricted variance covariance and the second model is with diagonal variance covariance (the correlation between the endogenous variables is explained entirely by the latent class variable). Here, neither of the two models is able to recover the latent classes well. This points out again the need for a small EFA style correlation structure to account for endogenous variables correlation not accounted for by the latent class variable. The EFA model is able to obtain parsimonious correlation structure while also retaining the correct latent class formation despite the moderate entropy level and relatively low class separation.

11. Alignment in Factor Mixture Analysis

Factor Mixture Analysis has been discussed in Lubke and Muthén (2005, 2007), Lubke et al. (2007), Clark et al. (2009, 2013). Similar to the previous example, a latent class model is estimated where the endogenous variables are not conditionally independent but within each class the dependence is modeled with a factor analysis model. In these factor mixture models, it is desirable to estimate the model with a class invariant loading matrix. This gives a more parsimonious model than a model with class specific loading structure. As a result, typically a better BIC can be obtained with such models. This is important because BIC is widely used for model comparison in mixture settings where often models are compared with different number of classes and factors. Invariant loading structure also allows us to estimate class specific factor variance for all non-reference classes. If the endogenous variables intercepts are also estimated as class invariant then the mean of the factor can be estimated. These class invariant restrictions however might not hold. We can use PSEM, however, to obtain a model with as much invariance as possible. This mixture model then becomes equivalent to the multiple group factor analysis alignment model, see Asparouhov and Muthén (2014b), with the only difference being that the subpopulation variable is not the observed grouping variable but is the unobserved latent class variable. Essentially, we suggest using alignment in factor mixture models as an alternative to using strict class invariance.

We illustrate this concept with a 2-class model where 8 continuous variables are used as class indicators. In addition, within each class the 8 variables measure two factors. Each factor loads on 5 variables. There are 3 non-invariant intercepts and 1 non-invariant loading. We use ALF DIFF priors for all intercepts and loadings to obtain the factor analysis alignment across the two classes. It should be noted here that the non-invariance parameters are helpful in identifying the latent classes. In the absence of non-invariance, the latent classes would be measured only by the latent class differences in the factor distribution, which is relatively small in the following sense. In LPA, 8 intercept differences contribute to the class formation. In this alignment of the factor mixture analysis model, only two factor intercepts and any other non-invariant intercept contribute to the latent class formation. Thus non-invariance in the intercepts of the observed variables is important in the class identification.

Figure 18 contains the input file for the simulation study. The model population is omitted for brevity but is identical to the model statement. Figure 19 contains the results of

```
                              ESTIMATES              S. E.     M. S. E.  95%  % Sig
                 Population   Average   Std. Dev.   Average              Cover Coeff
Latent Class 1
 F1        BY
  Y1               1.000     1.0115     0.0436     0.0468     0.0020 0.960 1.000
  Y2               1.000     1.0068     0.0366     0.0431     0.0014 0.980 1.000
  Y5               0.500     0.4981     0.0401     0.0474     0.0016 0.980 1.000
  Y6               0.000    -0.0006     0.0243     0.0289     0.0006 1.000 0.000
  Y7               0.000    -0.0033     0.0314     0.0317     0.0010 0.980 0.020
  Y8               0.000    -0.0022     0.0292     0.0320     0.0008 0.990 0.010
 F2        BY
  Y1               0.200     0.1823     0.0466     0.0466     0.0025 0.920 0.930
  Y2               0.000    -0.0187     0.0267     0.0313     0.0011 0.980 0.020
  Y5               0.500     0.4965     0.0421     0.0444     0.0018 0.970 1.000
  Y6               0.500     0.4962     0.0382     0.0386     0.0015 0.940 1.000
  Y7               0.500     0.5016     0.0434     0.0397     0.0019 0.920 1.000
  Y8               0.500     0.5020     0.0357     0.0398     0.0013 0.960 1.000
 F1          ON
  X                0.400     0.3931     0.0451     0.0456     0.0021 0.960 1.000
 F2          ON
  X               -0.200    -0.2063     0.0494     0.0565     0.0025 0.990 0.990
 F1        WITH
  F2               0.400     0.4129     0.0519     0.0597     0.0028 0.950 1.000

Latent Class 2
 F1        BY
  Y1               0.500     0.4971     0.0334     0.0406     0.0011 0.970 1.000
  Y2               0.500     0.4992     0.0381     0.0386     0.0014 0.940 1.000
  Y5               0.300     0.2868     0.0429     0.0464     0.0020 0.960 1.000
  Y6               0.000    -0.0109     0.0329     0.0308     0.0012 0.950 0.050
  Y7               0.000    -0.0135     0.0308     0.0305     0.0011 0.940 0.060
  Y8               0.000    -0.0049     0.0276     0.0297     0.0008 0.990 0.010
 F2        BY
  Y1               0.200     0.1892     0.0361     0.0399     0.0014 0.950 1.000
  Y2               0.000    -0.0068     0.0275     0.0265     0.0008 0.970 0.030
  Y5               1.000     0.9984     0.0435     0.0469     0.0019 0.980 1.000
  Y6               1.000     1.0040     0.0422     0.0426     0.0018 0.960 1.000
  Y7               1.000     1.0069     0.0409     0.0426     0.0017 0.970 1.000
  Y8               1.000     0.9973     0.0411     0.0429     0.0017 0.970 1.000
 F1          ON
  X                0.200     0.2031     0.0461     0.0501     0.0021 0.970 1.000
 F2          ON
  X                0.400     0.4034     0.0476     0.0477     0.0023 0.970 1.000
 F1        WITH
  F2               0.200     0.2271     0.0523     0.0614     0.0034 0.940 0.990
```

Figure 17. ESEM mixture results.

this simulation for a selection of the parameters. The bias in the parameter estimates is minimal and the coverage is near the nominal level. The alignment of the factor mixture analysis model is not the most parsimonious model that can be estimated in these settings. Ideally, this analysis should be followed by the estimation of a standard factor mixture analysis where parameters that are identified as non-invariant are estimated as class specific while parameters that were identified as invariant are held equal across class. Such a model would yield the most parsimonious and well fitting model and thus would yield the best BIC value.

Another model that is of similar interest is the model where alignment is performed only on the loadings but not on the means. The means of the factors will be fixed to zero in each class but the endogenous variable means will be class specific as in LPA. Such a model would have more power to identify homogeneous subgroups in the population because there are more class-specific parameters.

12. Latent Profile Analysis with PSEM

Latent profile analysis uses continuous variables as indicators for latent classes. The most common version estimates class specific means for the variables, assumes conditional independence within each class, and class invariant variance parameters. The conditional independence assumption is often unrealistic and within class some residual covariance occurs. It is possible to abandon the conditional independence and class invariance for the variance covariance and still estimate the LPA model, i.e., conditional on the latent class variable the distribution of the class-indicator variables is class-specific unconstrained multivariate

```
montecarlo:
 names are y1-y8;
 genclasses = c(2);
 classes = c(2);
 nobs = 2000;
 nrep = 100;

analysis: type = mixture;

model:
 %overall%
 f1 by y1-y4*1 y5*0.5;
 f2 by y5-y8*1 y1*0.2;
 y1-y8*.5;

 %c#1%
 f1 by y1-y4*1 y5*0.5 (a1-a5);
 f2 by y5-y8*.5 y1*0.2(a6-a10);
 [y1*2 y2*2 y3*2 y4*2 y5*-1 y6*-1 y7*-1 y8*-1] (m1-m8);
 f1 with f2*0.4;
 f1-f2@1;
 [f1-f2@0];

 %c#2%
 f1 by y1-y4*1 y5*0.2 (b1-b5);
 f2 by y5-y8*.5 y1*0.2 (b6-b10);
 [f1*1 f2*1];
 f1*0.8 f2*1.2;
 [y1*0 y2*2 y3*0 y4*2 y5*-1 y6*0 y7*-1 y8*-1] (n1-n8);
 f1 with f2*0.2;

model prior:
 DO(#,1,10) DIFF(a# b#)~ALF(0,1);
 DO(#,1,8) DIFF(m# n#)~ALF(0,1);
```

Figure 18. Alignment in factor mixture analysis.

normal distribution. This model can be identified with large samples but the number of additional parameters makes the model too flexible. With smaller samples the model will likely be poorly identified. In the previous two sections we discussed one possible parsimonious strategy to model the within class covariance: using within class factor analysis models. In this section we focus on the situation where factors are not easily available such as sparse covariance matrices. In such situations, the problem with parsimony can be resolved by adding ALF(0,v) priors for all covariance parameters, i.e., forcing these parameters to be as close to 0 as possible. Because the model without the priors is identified, this model is a RegSEM model and not a PSEM model.

We illustrate the advantages of PSEM-LPA with a simulation study using a 2-class model with 8 indicators. Within each class we add 3 non-zero covariances. We compare three models: the standard LPA with conditional independence and variance invariance, LPA with class-specific unconstrained variance/covariance matrix, and the LPA-PSEM model which adds the ALF(0,1) priors for all covariances. The PSEM-LPA simulation study is given in Figure 20. The results of the simulation study are given in Figure 21 where we compare the three models for a selection of the parameters. PSEM-LPA outperforms the two standard LPA models in terms of parameter bias, coverage and MSE. This also implies that the PSEM-LPA is better at recovering the latent subgroups.

13. Growth Modeling with Non-Normal Outcomes

The PSEM framework allows us to estimate more flexible growth models. Subject specific latent variables (intercept and slope) are used to model individually specific developmental curves. Typically, in a linear growth model, the means of the random intercept and slope are estimated but the intercepts of the observed variables are fixed to zero. If one attempts to estimate the intercepts of the observed variables as well, the model becomes unidentified. The PSEM framework allows us to estimate both the observed variable intercepts as well as the means of the random intercept and slope by giving LASSO or ALF priors to the observed variable intercepts. That allows us to estimate the random intercept and slope means while still allowing small time-specific deviations from the linear growth projection when such are needed. These time-specific deviations occur quite often and usually yield substantial improvement in data fit.

This model is discussed in detail in Section 4.2 in Asparouhov and Muthén (2024) for continuous outcomes. Here we will illustrate the model for count outcomes using the PML estimator in conjunction with numerical integration which is required for the estimation with count outcomes. The growth model is as follows. The observed count variable for individual $i = 1, \ldots, N$ at time $t = 0, 1, \ldots, T$ is

$$Y_{it} \sim Po(\mu_{it}) \tag{22}$$

where Po here represents the Poisson distribution. Furthermore,

$$\log(\mu_{it}) = \nu_t + I_i + S_i \cdot t \tag{23}$$

where ν_t is the time specific deviation, I_i and S_i are the random intercepts and slope for individual i and

$$\begin{pmatrix} I_i \\ S_i \end{pmatrix} \sim N(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}) \tag{24}$$

$$\nu_t \sim ALF(0, 1). \tag{25}$$

As usual, the ALF prior/penalty for ν_t will minimize these parameters near zero and will minimize how many of them are non-zero.

The above model uses two-dimension integration. In our illustration, we fix σ_{12} and σ_{22} to zero, which results in one dimensional integration and faster computations. In this simulation study, we use $T = 7$, i.e., the growth model is based on 8 count observations. Two of the time points do not fit the linear growth model perfectly, i.e., two of the ν_t parameters are non-zero. The PSEM model estimation is able to detect which two of the time points need time specific parameters ν_t while at the same time, the distribution of the random effects is estimated without bias. The simulation study setup is given in Figure 22 and the results are given in Figure 23. The bias is minimal and the coverage for the estimates is near the nominal level.

This model generalizes to various other link functions and outcome variables, for example censored outcomes, negative binomial, and ordered categorical. For categorical variables, the model can be estimated with the WLSMV estimator or the ML estimator with numerical integration. In the presence of missing data, the ML estimator as usual has

```
                              ESTIMATES              S. E.     M. S. E.  95%  % Sig
                 Population   Average   Std. Dev.   Average              Cover Coeff

 Latent Class 1
  F1       BY
   Y1                1.000     0.9968     0.0426     0.0385     0.0018 0.920 1.000
   Y2                1.000     0.9952     0.0325     0.0340     0.0011 0.960 1.000
   Y5                0.500     0.5033     0.0410     0.0356     0.0017 0.890 1.000
  F2       BY
   Y5                0.500     0.4932     0.0362     0.0373     0.0013 0.950 1.000
   Y6                0.500     0.4945     0.0336     0.0321     0.0011 0.950 1.000
   Y1                0.200     0.2006     0.0366     0.0387     0.0013 0.960 1.000
  F2       WITH
   F1                0.400     0.3968     0.0446     0.0421     0.0020 0.930 1.000
  Intercepts
   Y1                2.000     1.9983     0.0447     0.0437     0.0020 0.930 1.000
   Y2                2.000     1.9966     0.0426     0.0413     0.0018 0.950 1.000
   Y3                2.000     2.0011     0.0424     0.0413     0.0018 0.940 1.000

 Latent Class 2
  F1       BY
   Y1                1.000     0.9951     0.0413     0.0412     0.0017 0.940 1.000
   Y2                1.000     1.0043     0.0403     0.0398     0.0016 0.970 1.000
   Y5                0.200     0.1953     0.0340     0.0343     0.0012 0.950 1.000
  F2       BY
   Y5                0.500     0.4974     0.0376     0.0343     0.0014 0.960 1.000
   Y8                0.500     0.4955     0.0303     0.0339     0.0009 0.980 1.000
   Y1                0.200     0.1988     0.0309     0.0316     0.0009 0.960 1.000
  F2       WITH
   F1                0.200     0.1967     0.0492     0.0479     0.0024 0.960 1.000
  Means
   F1                1.000     0.9941     0.0616     0.0695     0.0038 0.970 1.000
   F2                1.000     1.0296     0.0856     0.1031     0.0081 0.990 1.000
  Intercepts
   Y1                0.000     0.0113     0.0749     0.0748     0.0057 0.960 0.040
   Y2                2.000     2.0080     0.0493     0.0484     0.0025 0.960 1.000
   Y3                0.000     0.0135     0.0659     0.0659     0.0045 0.930 0.070
  Variances
   F1                0.800     0.8055     0.0708     0.0673     0.0050 0.940 1.000
   F2                1.200     1.2371     0.1367     0.1513     0.0199 0.960 1.000
```

Figure 19. Alignment in factor mixture analysis results.

the advantage that the parameter estimates are unbiased even when the missing data is MAR.

14. EFA with Nominal Variables

Exploratory factor analysis can be estimated in Mplus with continuous, ordered categorical, censored and count variables. Using the PSEM methodology we can now add to this list nominal variables. Numerical integration is used for the estimation of such models. For models with a larger number of factors, the numerical integration can be performed with fewer integration points per dimension. Alternatively, Monte Carlo integration can be used. With nominal variables, the interpretation of the EFA model is somewhat more complex because one nominal variable with K categories will have $K-1$ loadings. If N represents the nominal variable, η represents the vector of latent variables, the model for N is given by

$$P(N=k)=\frac{Exp(\nu_k+\lambda_k\eta)}{\sum_{k=1}^{K}Exp(\nu_k+\lambda_k\eta)} \tag{26}$$

The last loading vector λ_K is fixed to 0 for identification purposes as well as ν_K. If a nominal variable is connected to a particular factor we can expect all $K-1$ loadings to be non-zero. If a nominal variable is not connected to a factor we can expect all $K-1$ loadings to be zero. If some of the $K-1$ loadings are zero and some are not zero, the interpretation is somewhat more difficult because the zeros and non-zeros change depending on which nominal category is set as the reference category. Thus, one key question in this kind of analysis is whether or not the EFA model is affected by the reference category of the nominal variables. The reference category for the nominal variables is set arbitrarily since the categories are not ordered and it would be important for the EFA estimation to not be affected by this arbitrary setting.

```
montecarlo:
 names are y1-y8;
 genclasses = c(2);
 classes = c(2);
 nobs = 300;
 nrep = 100;

model population:

 %overall%
 y1-y8*8;

 %c#1%
 [y1-y8*1];
 y1 with y5*1;
 y2 with y6*1;
 y3 with y7*1;

 %c#2%
 [y1-y8*-1];
 y2 with y4*3;
 y1 with y4*3;
 y1 with y2*3;

analysis: type = mixture;

model:
 %overall%
 y1-y8*8;

 %c#1%
 [y1-y8*1];
 y1-y8*8;
 y1-y8 with y1-y8*0 (p1-p28);

 %c#2%
 [y1-y8*-1];
 y1-y8*8;
 y1-y8 with y1-y8*0 (q1-q28);

model prior:
 p1-p28~ALF(0,1);
 q1-q28~ALF(0,1);
```

Figure 20. PSEM-LPA simulation study.

For the PSEM estimation of EFA models with nominal variables we describe here, the rotation criterion is applied to all the loadings as if they come from different variables. An alternative method might be possible that takes into account the fact that the $K-1$ loadings are intra-connected and that the loadings can change if the reference category is altered. Ultimately, a rotation criterion for the nominal variable should really be symmetric with respect to which category is selected as a reference category. One possible criterion is as follows: for each nominal variable create $K(K-1)$ rows in the loadings matrix which consists of the $K-1$ loadings crossed with the K possible reference categories. If $K=3$ and the loading vectors with the last category as a reference category are λ_1 and λ_2, the total loading matrix will have these 6 rows λ_1, λ_2, $\lambda_1-\lambda_2$, $-\lambda_2$, $\lambda_2-\lambda_1$, $-\lambda_1$. This rotation criterion is then invariant to the reference category but it is not necessarily optimal since we want the loading structure to be simple for one of the reference categories and not all of them. We will not explore this here and will use for simplicity only the $K-1$ loading rows. Further research is needed in this direction, however, we want to point out here again that this concept is of concern only for the case when the $K-1$ loadings of a nominal variable on a factor includes both zeros and non-zeros. If the number of nominal variables of that kind is relatively small as compared to the total number of observed variables in the EFA model, it is unlikely that a more elaborate rotation criterion will lead to a different solution.

Consider also the following related aspect of the EFA analysis. If a nominal variable has zero and non-zero loadings for the same factor, it may be possible to exclude that variable from the rotation. If there is a sufficient number of other observed variables that can be used to identify the EFA model, this kind of a nominal variable can be used as a distal outcome, i.e., the loadings are estimated as well but they are excluded from the rotation criterion completely. This situation may indeed be preferable in the case when the nominal variable is truly a distal outcome that is regressed on the EFA factors, rather than a part of the measurement model.

As stated earlier, these complications arise in some special situations and may not need to be dwelled on in many simpler examples. In this section we use as an illustration an EFA model with 5 continuous indicators, 3 nominal variables, and 2 factors. The nominal variables have 3 categories. Of the 3 nominal variables 2 are pure indicators, i.e., for two of the nominal variables both loadings are simultaneously zero or simultaneously non-zero for each of the two factors. The third nominal variable has zero and non-zero loadings on the same factor. The full setup for this simulation study is given in Figure 24. We use all the loadings for the rotation. The results of the simulation study are given in Figure 25. The bias is minimal and the coverage is near the nominal levels. Using only the continuous and the 2 pure nominal indicators to determine the optimal rotation yields nearly identical results.

Other types of variables can similarly be used in an EFA model. For example time-to-even/survival variables can now also be used as indicators in EFA models.

15. Measurement Invariance across Levels in Multi-Level Factor Analysis

Consider the following two-level factor analysis model. Let Y_{ijp} be the p−th observed variable for individual i in cluster j. The variables measure one factor on the within level η_{wij} and one factor on the between level η_{bj}

$$Y_{ijp} = Y_{wijp} + Y_{bjp} \tag{27}$$

$$Y_{wijp} = \lambda_{wp}\eta_{wij} + \varepsilon_{wijp} \tag{28}$$

$$Y_{bjp} = \nu_p + \lambda_{bp}\eta_{bj} + \varepsilon_{bjp} \tag{29}$$

$$\eta_{wij} \sim N(0,\psi_w), \eta_{bj} \sim N(0,\psi_b), \varepsilon_{wijp} \sim N(0,\theta_{wp}), \varepsilon_{bjp} \sim N(0,\theta_{bp}). \tag{30}$$

Certain constraints are required to identify the model. One common approach is to fix the factor variances ψ_w and ψ_b to 1. In most applications, it is desirable to interpret the two factors as being the within and the between part of the same latent construct. Such interpretation requires the

```
                                  ESTIMATES              S. E.     M. S. E.  95%  % Sig
                 Population   Average    Std. Dev.   Average              Cover Coeff

 LPA with conditional independence
 Latent Class 1
  Means
   Y1                 1.000     1.1561     0.3182     0.2749     0.1246 0.850 0.980
   Y2                 1.000     1.1529     0.2717     0.2776     0.0965 0.900 0.990
   Y3                 1.000     0.7174     0.3530     0.3230     0.2032 0.770 0.600
 Latent Class 2
  Means
   Y1                -1.000    -1.7712     0.4428     0.4483     0.7888 0.620 0.960
   Y2                -1.000    -1.7591     0.4712     0.4513     0.7960 0.620 0.990
   Y3                -1.000    -1.0246     0.3121     0.3365     0.0970 0.950 0.870

 LPA with full covariance
 Latent Class 1
  Means
   Y1                 1.000     0.8409     0.5891     0.4298     0.3689 0.780 0.550
   Y2                 1.000     0.8814     0.5972     0.4736     0.3672 0.770 0.570
   Y3                 1.000     0.8834     0.6482     0.5070     0.4296 0.760 0.530
 Latent Class 2
  Means
   Y1                -1.000    -0.9290     0.5929     0.5365     0.3531 0.740 0.570
   Y2                -1.000    -1.0239     0.5953     0.5482     0.3514 0.790 0.520
   Y3                -1.000    -0.8995     0.6154     0.5139     0.3850 0.780 0.460

 PSEM-LPA
 Latent Class 1
  Means
   Y1                 1.000     1.0791     0.3876     0.3146     0.1550 0.900 0.940
   Y2                 1.000     1.0986     0.3163     0.3234     0.1088 0.940 0.950
   Y3                 1.000     0.9918     0.3768     0.3612     0.1406 0.900 0.830
 Latent Class 2
  Means
   Y1                -1.000    -1.1693     0.4210     0.3779     0.2042 0.940 0.850
   Y2                -1.000    -1.2106     0.4365     0.3844     0.2330 0.880 0.840
   Y3                -1.000    -1.0439     0.3740     0.3383     0.1404 0.890 0.860
```

Figure 21. PSEM-LPA simulation study results.

```
montecarlo:
        names are u1-u8;
        generate = u1-u8(c);
        count = u1-u8;
        nobs = 2000;
        nrep = 100;

analysis: algorithm = integration;

model population:
        i s | u1@0 u2@1 u3@2 u4@3 u5@4 u6@5 u7@6 u8@7;
        i*.7; s@0; i with s@0;
        [i*-2 s*0.6];
        [u1-u3*0 u4*0.4 u5-u7*0 u8*-0.3];

model:
        i s | u1@0 u2@1 u3@2 u4@3 u5@4 u6@5 u7@6 u8@7;
        i*.7; s@0; i with s@0;
        [i*-2 s*0.6];
        [u1-u3*0 u4*0.4 u5-u7*0 u8*-0.3] (m1-m8);

model prior: m1-m8~ALF(0,1);
```

Figure 22. PSEM growth model for counts simulation study.

following formulation of the factor model

$$Y_{ijp} = \nu_p + \lambda_p \eta_{ij} + \varepsilon_{wijp} + \varepsilon_{bjp} \tag{31}$$

$$\eta_{ij} = \eta_{wij} + \eta_{bj} \tag{32}$$

$$\eta_{wij} \sim N(0, \psi_w), \eta_{bj} \sim N(0, \psi_b), \varepsilon_{wijp} \sim N(0, \theta_{wp}), \varepsilon_{bjp} \sim N(0, \theta_{bp}). \tag{33}$$

Typically, this model is identified by fixing the within level factor variance ψ_w to 1. Equation (32) allows us to interpret the between part of the factor as the random intercept as in two-level random intercept regression. The between part of the factor is the cluster level contribution to the latent construct while the within factor is the individually specific part of the latent construct. The intra class correlation for the factor is computed as

$$ICC_\eta = \frac{\psi_b}{1 + \psi_b} \tag{34}$$

```
                                 ESTIMATES              S. E.     M. S. E.  95%  % Sig
                Population   Average   Std. Dev.    Average              Cover Coeff
Means
 I                  -2.000     -1.9871    0.0338     0.0394      0.0013 0.980 1.000
 S                   0.600      0.5964    0.0075     0.0069      0.0001 0.910 1.000
Intercepts
 U1                  0.000     -0.0066    0.0331     0.0349      0.0011 1.000 0.000
 U2                  0.000     -0.0134    0.0294     0.0330      0.0010 0.980 0.020
 U3                  0.000     -0.0032    0.0318     0.0281      0.0010 0.970 0.030
 U4                  0.400      0.3970    0.0231     0.0227      0.0005 0.950 1.000
 U5                  0.000      0.0030    0.0162     0.0146      0.0003 0.980 0.020
 U6                  0.000      0.0053    0.0169     0.0097      0.0003 0.950 0.050
 U7                  0.000      0.0082    0.0192     0.0111      0.0004 0.970 0.030
 U8                 -0.300     -0.2887    0.0268     0.0187      0.0008 0.950 0.990
Variances
 I                   0.700      0.6979    0.0269     0.0252      0.0007 0.910 1.000
```

Figure 23. PSEM growth model for counts results.

```
montecarlo:
   names are u1-u8;
   generate = u6-u8(n 2);
   nominal = u6-u8;
   nobs = 2000;
   nrep = 100;

analysis: algorithm = integration;

model population:
f1 by u1-u4*1 u5*0 u6#1*0 u6#2*0 u7#1*0 u7#2*0 u8#1*0 u8#2*0.4;
f2 by u1*0.3 u2-u4*0 u5*0.8 u6#1*0.5 u6#2*1 u7#1*1 u7#2*0.5 u8#1*0.4 u8#2*1;
f1 with f2*0.4; f1-f2@1;
[u1-u3*0 u4*0.4 u5*0 u6#1*0.5 u6#2*-0.5 u7#1*.3 u7#2*0.5 u8#1*0.4 u8#2*0];
u1-u5*1;

model:
f1 by u1-u4*1 u5*0 u6#1*0 u6#2*0 u7#1*0 u7#2*0 u8#1*0 u8#2*0.4 (a1-a11);
f2 by u1*0.3 u2-u4*0 u5*0.8 u6#1*0.5 u6#2*1 u7#1*1 u7#2*0.5 u8#1*0.4 u8#2*1 (a12-a22);
f1 with f2*0.4; f1-f2@1;
[u1-u3*0 u4*0.4 u5*0 u6#1*0.5 u6#2*-0.5 u7#1*.3 u7#2*0.5 u8#1*0.4 u8#2*0];
u1-u5*1;

model prior: a1-a22~geomin(2,1);
```

Figure 24. EFA with nominal variables.

In the above model and model (27–30), the term ε_{bjp} is often omitted, i.e., θ_{bp} is fixed to 0. The following discussion applies equally well with or without that term.

Model (31–33) is nested within the model (27–30). The models are equivalent if the within level loadings and the between level loadings in (27–30) are proportional (across indicators)

$$\lambda_{wp} = \lambda_p \tag{35}$$

$$\lambda_{bp} = \sqrt{\psi_b}\lambda_p \tag{36}$$

where $\sqrt{\psi_b}$ is the coefficient of proportionality. If this proportionality condition holds then model (27–30) can be rewritten as

$$Y_{ijp} = Y_{wijp} + Y_{bjp} \tag{37}$$

$$Y_{wijp} = \lambda_p \eta_{wij} + \varepsilon_{wijp} \tag{38}$$

$$Y_{bjp} = \nu_p + \lambda_p \eta_{bj} + \varepsilon_{bjp} \tag{39}$$

$$\eta_{wij} \sim N(0,1), \eta_{bj} \sim N(0,\psi_b), \varepsilon_{wijp} \sim N(0,\theta_{wp}), \varepsilon_{bjp} \sim N(0,\theta_{bp}). \tag{40}$$

Model (37–40) is identical to the model (31–33). This proportionality condition can also be viewed as measurement invariance across the two levels, similarly to how measurement invariance is considered in multiple group factor analysis.

It is often the case that the proportionality condition is violated for some of the factor indicators, i.e., only a partial invariance holds across the levels. Thus we are faced with the dilemma to use a poorly fitting model (37–40) or a better fitting model (27–30) but without the interpretability of the factor as the within-between parts of the same latent construct. The PSEM framework can be used to resolve this problem, similarly to multiple group alignment. We estimate the model (27–30) with

```
                                 ESTIMATES                S. E.     M. S. E.  95%  % Sig
                    Population   Average   Std. Dev.   Average               Cover Coeff

 F1         BY
  U1                     1.000     1.0056     0.0312     0.0452     0.0010 0.990 1.000
  U2                     1.000     0.9978     0.0345     0.0368     0.0012 0.970 1.000
  U3                     1.000     1.0071     0.0312     0.0368     0.0010 1.000 1.000
  U4                     1.000     0.9957     0.0332     0.0365     0.0011 0.970 1.000
  U5                     0.000    -0.0031     0.0394     0.0668     0.0015 0.980 0.020
  U6#1                   0.000    -0.0036     0.0689     0.0698     0.0047 0.980 0.020
  U6#2                   0.000     0.0028     0.0904     0.0845     0.0081 0.970 0.030
  U7#1                   0.000    -0.0021     0.0716     0.0759     0.0051 0.960 0.040
  U7#2                   0.000    -0.0056     0.0695     0.0731     0.0048 0.950 0.050
  U8#1                   0.000    -0.0215     0.0805     0.0822     0.0069 0.960 0.040
  U8#2                   0.400     0.3780     0.0948     0.1237     0.0094 0.990 0.870

 F2         BY
  U1                     0.300     0.2903     0.0362     0.0540     0.0014 0.980 1.000
  U2                     0.000     0.0041     0.0372     0.0396     0.0014 0.970 0.030
  U3                     0.000    -0.0106     0.0341     0.0393     0.0013 0.990 0.010
  U4                     0.000    -0.0035     0.0300     0.0388     0.0009 0.990 0.010
  U5                     0.800     0.8044     0.0629     0.0722     0.0039 0.980 1.000
  U6#1                   0.500     0.4863     0.0953     0.0994     0.0092 0.980 1.000
  U6#2                   1.000     0.9813     0.1437     0.1364     0.0208 0.930 1.000
  U7#1                   1.000     0.9926     0.1329     0.1293     0.0175 0.950 1.000
  U7#2                   0.500     0.5002     0.1033     0.1114     0.0106 0.960 1.000
  U8#1                   0.400     0.4050     0.1115     0.1101     0.0123 0.980 0.980
  U8#2                   1.000     1.0269     0.1716     0.1527     0.0299 0.920 1.000

 F1         WITH
  F2                     0.400     0.4063     0.0220     0.0659     0.0005 1.000 0.990
```

Figure 25. EFA with nominal variables results.

ψ_b as a free parameter. This of course is an unidentified model on its own. Instead of holding the loadings equal across the two levels, as in (37–40), we use approximate measurement invariance as in alignment. That is, we add ALF priors for $\lambda_{wp} - \lambda_{bp}$. This will allow us to obtain a model with the same fit as model (27–30), while also obtaining the closest model to measurement invariance, and as a result, an approximate within-between interpretation of the factor.

We illustrate this PSEM estimation for a two-level factor model with 4 indicators measuring 1 factor on both levels. In this simulation study, only one of the four indicators lacks measurement invariance. We generate and analyze samples with 200 clusters of size 30. Figure 26 gives the input file for this simulation study. The results are given in Figure 27. The bias is small and the coverage is near the nominal levels. The non-invariant indicator is correctly identified. The invariant indicator loadings are nearly identical across the two levels. The average chi-square value for this model is 3.1 and with 4 degrees of freedom we obtain a rejection rate of 1%.

The measurement invariance across levels PSEM model applies also to 3-level models and models with other types of indicators such as categorical indicators. The model can be estimated with the ML and WLSMV estimators in Mplus.

16. Multilevel Exploratory Structural Equation Models

In this section, we combine multilevel models and exploratory factor analysis models with the power of the PSEM framework. Let Y_{ij} be a vector of observed variables for individual i in cluster j. Let η_{ij} be a vector of exploratory factors measured by Y_{ij}. Let X_{ij} be a vector of covariates. The two-level ESEM model is given by

$$Y_{ij} = \nu_j + \Lambda\eta_{ij} + \varepsilon_{ij} \quad (41)$$

$$\eta_{ij} = \alpha_j + B_j X_{ij} + \xi_{ij}. \quad (42)$$

Here ν_j is a vector of random intercepts for the observed variables, Λ is an unconstrained matrix of loadings to be rotated using a simplicity Geomin rotation criterion, ε_{ij} is a vector of uncorrelated residuals, α_j is a vector of random intercepts for the EFA factors. For identification purposes α_j means are fixed to 0. The matrix B_j contains cluster-specific random regression coefficients. ξ_{ij} is a vector of correlated residuals with variance fixed to 1 for identification purposes. Typically α_j and B_j are fully correlated random effects but ν_j are uncorrelated random effects (among each other and also other effects). The reason for this is that α_j acts as between level factors and ν_j as residuals on the between level. In fact, α_j is the vector of between level factors which is added to the within level factors to obtain the total factors η_{ij}. The

```
montecarlo:
        names are y1-y4;
        nobservations = 6000;
        ncsizes = 1;
        csizes = 200(30);
        nreps = 100;

ANALYSIS: TYPE IS TWOLEVEL; tolerance=0.0001;

MODEL POPULATION:

        %Within%
        fw BY y1-y4*1;
        y1-y4*1;
        fw@1;

        %Between%
        fb BY y1*1 y2*1.4 y3*1 y4*1;
        fb*.4;
        y1-y4*0.5;

MODEL:

        %Within%
        fw BY y1-y4*1 (a1-a4);
        y1-y4*1;
        fw@1;

        %Between%
        fb BY y1*1 y2*1.4 y3*1 y4*1 (b1-b4);
        fb*.4;
        y1-y4*0.5;


model prior: do(#,1,4) DIFF(a# b#)~ALF(0,1);
```

Figure 26. Measurement invariance in two-level factor analysis.

same loading matrix applies to the within and the between portions of the factors. Thus, this model is based on across-level measurement invariance which was discussed in the previous section.

We present this model for normally distributed outcomes but the model generalizes to other types of variables if numerical integration is utilized. The model can be modified and generalized in various ways: direct effects from the covariates can be included, ν_j can be non-random intercepts, etc. The model essentially is an EFA model in Equation (41) and a standard hierarchical model in Equation (42) where the EFA factors are now the dependent variables.

We illustrate the multilevel ESEM model with a stimulation study using a 2-factor EFA model, measured by 10 indicator variables, and one covariate. In the EFA model, each factor has 5 main indicators as well as one cross-loading. The simulation study setup is given in Figure 28. The random intercepts for the factors are estimated as factors on the between level. The loadings are held equal across the two levels to ensure that the between factors indeed take the role of random intercepts. The results of the simulation study for a selection of the parameters are given in Figure 29. The bias in the parameter estimates is minimal and the coverage is near the nominal level. The Geomin penalty function responsible for the factor rotation successfully rotates the factors to the simplest loading pattern and the external multilevel model as well. The estimation of this model is very fast: it takes only 16 seconds to estimate 100 replications.

17. Lasso Regression

Penalized methods were originally developed for linear regression models for the situation where a variable is regressed on

	Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Within Level							
FW BY							
Y1	1.000	0.9976	0.0186	0.0183	0.0003	0.930	1.000
Y2	1.000	0.9988	0.0197	0.0181	0.0004	0.940	1.000
Y3	1.000	0.9993	0.0175	0.0182	0.0003	0.970	1.000
Y4	1.000	0.9982	0.0183	0.0180	0.0003	0.950	1.000
Variances							
FW	1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
Between Level							
FB BY							
Y1	1.000	0.9825	0.0894	0.0955	0.0082	0.960	1.000
Y2	1.400	1.3566	0.1665	0.1897	0.0293	0.930	1.000
Y3	1.000	0.9693	0.0876	0.0813	0.0085	0.960	1.000
Y4	1.000	0.9752	0.1158	0.0982	0.0139	0.970	1.000
Variances							
FB	0.400	0.4191	0.0703	0.0945	0.0053	0.980	1.000

Figure 27. Measurement invariance in two-level factor analysis results.

```
montecarlo:
        names are y1-y10 x;
        nobservations = 6000;
        ncsizes = 1;
        csizes = 200(30);
        nreps = 100;
        within=x;

ANALYSIS: type=twolevel random;

MODEL POPULATION:

        %Within%
        f1 BY y1-y5*1 y6*0.3 y7-y10*0 (a1-a10);
        f2 BY y1-y4*0 y5*0.5 y6-y10*0.7 (a11-a20);
        y1-y10*1; f1-f2@1; f1 with f2*0.3; x*1;
        s1 | f1 on x;
        s2 | f2 on x;

        %Between%
        f1b BY y1-y5*1 y6*0.3 y7-y10*0 (a1-a10);
        f2b BY y1-y4*0 y5*0.5 y6-y10*0.7 (a11-a20);
        f1b-f2b*0.6; f1b with f2b*0.3; y1-y10*0.5;
        s1-s2*0.1; [s1*0.3 s2*0.5];

MODEL:

        %Within%
        f1 BY y1-y5*1 y6*0.3 y7-y10*0 (a1-a10);
        f2 BY y1-y4*0 y5*0.5 y6-y10*0.7 (a11-a20);
        y1-y10*1; f1-f2@1; f1 with f2*0.3;
        s1 | f1 on x;
        s2 | f2 on x;

        %Between%
        f1b BY y1-y5*1 y6*0.3 y7-y10*0 (a1-a10);
        f2b BY y1-y4*0 y5*0.5 y6-y10*0.7 (a11-a20);
        f1b-f2b*0.6; f1b with f2b*0.3; y1-y10*0.5;
        s1-s2*0.1; [s1*0.3 s2*0.5];

model prior: a1-a20~geomin(2,1,.0001);
```

Figure 28. Multilevel ESEM.

a large set of predictors. If the number of observations in the data set is larger than the number of covariates the model is identified and thus it is categorized as a RegSEM model. For such models, the weight of the penalty cannot be determined numerically and there is an extensive methodology developed already on the optimal selection of the penalty weight for linear regression. Among the most popular methods are the adaptive lasso method, cross-validation, and minimizing AIC/BIC values. Only the last method is easily accessible in Mplus at this time. Here we will briefly discuss Mplus specifics related to the last method and will also include a chi-square test of fit for model selection.

Suppose that a Y variable is regressed on a set of covariates $X_1, ..., X_Q$. The LASSO regression estimates the model using $LASSO(0, v)$ prior for the Q regression coefficients. To be able to make a proper model selection, this model is evaluated on a grid of v values. For example, using $v = 0.1, 0.2, ..., 1$, we obtain 10 models. Some of the β_q coefficients will be shrunk to near zero and some will not. Here we need to decide which coefficients are considered 0 and which are not. There are several different criteria that can be used for this purpose. We can use statistical significance to see which coefficients must be retained. Alternatively, a low threshold value can be set. If all the predictors and the dependent variable are standardized, coefficients which are smaller than 0.1 (or 0.05) may be considered substantively unimportant as they have less than 1% contribution to the R^2 of the dependent variable. Next, a parallel set of models must be estimated where the coefficients that are considered zero are actually fixed to 0. Among the 10 models then a selection criteria such as BIC or AIC can be used to make the final selection model or alternatively the most parsimonious model that is not rejected by the chi-square test of fit can be used as the final model. It should be noted here that the BIC and AIC criterions in the LASSO runs can not be used for model selection because Mplus does not adjust the number of parameters when the model is RegSEM. Mplus will adjust the number of parameters only for a proper PSEM model. Thus the BIC criterion should be obtained only in the parallel run where the coefficients that are near zero in LASSO are actually fixed to 0.

Consider the following simulated example where a dependent variable is regressed on 10 standardized covariates, 7 of the coefficients are zero and 3 are not zero: 0.3, 0.2 and 0.1. The sample size is only $N = 100$ and the correlation between all the covariates is set to 0.8. In such difficult settings, significance for the coefficients is out of reach. Figure 30 shows the input file for estimating the LASSO regression using $LASSO(0, v)$ penalty with $v = 1$. We estimate this model on the following grid of v values: 0.05,0.1,0.2,0.3, … ,0.9,1.0. For v between 0.4 and 1.0, there are 6 coefficients greater than 0.1 by absolute values, for the 6 variables X_1, X_2, X_3, X_4, X_7, and X_{10}. For $v = 0.3$ there are 5 coefficients that remain above 0.1: X_1, X_2, X_3, X_4, X_7. For $v = 0.2$ there are 3 coefficients that are above 0.1: X_1, X_3, X_4. For $v = 0.1$ only X_1 and X_3 have coefficients above 0.1. For $v = 0.05$ there are no coefficients that are above 0.1. Neither of these models is the true model which includes only X_1, X_2 and X_3 as predictors. Next we evaluate the above standard regression models with $Q_0 = 6,5,3,2$, and 0 predictors, i.e., $Q - Q_0$ of the parameters are fixed to 0. The first 5 of these are not rejected by the chi-square test of fit. Only the model with $Q_0 = 0$ is rejected. Therefore, using the chi-square test of fit as the selection criterion, we select the most parsimonious model which includes only X_1 and X_3 as predictors. Using the BIC criterion the same model is identified as the best model.

The covariate X_2 was omitted in the selection process for this one replication. In these settings, even when X_1 and X_3 are the only covariates selected in the model, significance for the regression coefficients could not be established. If the simulation is repeated with a larger sample size, we expect that all 3 predictors will be correctly identified. Note, however, that even if the model with the true covariates X_1, X_2 and X_3 was included in the model comparison, the model with X_1 and X_3 alone would have been the final selection model since it is more parsimonious and is not rejected by the test of fit for this particular data set.

```
                              ESTIMATES              S. E.     M. S. E.  95%  % Sig
                 Population   Average   Std. Dev.   Average              Cover Coeff

 Within Level

  F1       BY
   Y1                 1.000     0.9988     0.0192     0.0188     0.0004 0.940 1.000
   Y2                 1.000     0.9989     0.0162     0.0189     0.0003 0.980 1.000
   Y3                 1.000     1.0008     0.0174     0.0190     0.0003 0.950 1.000
   Y4                 1.000     1.0012     0.0165     0.0188     0.0003 0.980 1.000
   Y5                 1.000     1.0035     0.0401     0.0445     0.0016 0.960 1.000
   Y6                 0.300     0.3022     0.0534     0.0589     0.0028 0.970 0.990
   Y7                 0.000     0.0018     0.0533     0.0588     0.0028 0.990 0.010
   Y8                 0.000     0.0028     0.0531     0.0588     0.0028 0.990 0.010
   Y9                 0.000     0.0004     0.0554     0.0590     0.0030 0.950 0.050
   Y10                0.000     0.0053     0.0551     0.0587     0.0030 0.950 0.050

  F2       BY
   Y1                 0.000     0.0007     0.0178     0.0237     0.0003 0.970 0.030
   Y2                 0.000     0.0004     0.0170     0.0236     0.0003 1.000 0.000
   Y3                 0.000    -0.0001     0.0183     0.0239     0.0003 0.990 0.010
   Y4                 0.000    -0.0032     0.0187     0.0240     0.0004 0.990 0.010
   Y5                 0.500     0.5022     0.0179     0.0287     0.0003 0.990 0.990
   Y6                 0.700     0.7041     0.0225     0.0247     0.0005 0.960 1.000
   Y7                 0.700     0.7026     0.0237     0.0238     0.0006 0.970 1.000
   Y8                 0.700     0.7001     0.0240     0.0236     0.0006 0.970 1.000
   Y9                 0.700     0.7046     0.0250     0.0238     0.0006 0.960 1.000
   Y10                0.700     0.7006     0.0246     0.0238     0.0006 0.940 1.000

  F1       WITH
   F2                 0.300     0.2926     0.0786     0.0824     0.0062 0.940 0.890

 Between Level

  F1B      WITH
   F2B                0.300     0.3005     0.0726     0.0756     0.0052 0.960 0.970

  Means
   S1                 0.300     0.2967     0.0248     0.0292     0.0006 0.950 1.000
   S2                 0.500     0.5022     0.0333     0.0312     0.0011 0.930 1.000

  Variances
   S1                 0.100     0.1004     0.0154     0.0146     0.0002 0.910 1.000
   S2                 0.100     0.0964     0.0162     0.0157     0.0003 0.900 1.000
   F1B                0.600     0.6047     0.0762     0.0758     0.0058 0.930 1.000
   F2B                0.600     0.6102     0.0866     0.0884     0.0075 0.930 1.000
```

Figure 29. Multilevel ESEM results.

18. Regularization for Moderated Nonlinear Factor Analysis (MNLFA)

The MNLFA model was introduced in Bauer and Hussong (2009). The model is a generalization of the MIMIC model and it allows covariates to affect the factor variance in addition to the factor mean. For simplicity, in this section we shall limit the discussion to a 1-factor model but models with multiple factors can be used as well. The 1-factor MNLFA model is given as follows

$$Y = \nu + \Lambda\eta + \varepsilon \tag{43}$$

$$\eta = B_1 X + \xi \tag{44}$$

$$\varepsilon \sim N(0, \Theta), \xi \sim N(0, \mathrm{Exp}(B_2 X)), \tag{45}$$

where Y is a vector of factor measurements, X is a vector of factor predictors, η is the factor, and ε is a vector of uncorrelated residuals. Unlike the MIMIC model, the residual factor variance is $\mathrm{Exp}(B_2 X)$ and varies across covariates. The model can be viewed also as a generalization of the multiple group scalar factor model in the following sense. If the covariate X is an unordered categorical variable, in the form of dummy covariates for each group, the model reduces precisely to the multiple group factor analysis scalar model where each group except the reference group has a group-specific factor mean and variance parameters. Thus, we can view the MNLFA as a continuum generalization of the multiple group factor analysis model. If the covariates X represent a continuum of background

```
MONTECARLO:
    NAMES = y x1-x10;
    NOBSERVATIONS = 100;
    NREPS = 1;

MODEL POPULATION:
    y on x1*0.3 x2*0.2 x3*0.1 x4-x10*0;
    x1-x10 with x1-x10*0.8;
    x1-x10*1; y*1;

MODEL:
    y on x1*0.3 x2*0.2 x3*0.1 x4-x10*0 (a1-a10);

ANALYSIS: tolerance=0.00001;

MODEL PRIORS:
a1-a10~LASSO(0,1);
```

Figure 30. Lasso regression.

variables and we want to determine how the factor mean and variance change over that continuum, we can use the MNLFA model. The MNLFA model can be used with categorical or continuous indicators.

The regularization of the MNLFA model has been discussed in detail in Belzak and Bauer (2024). From a PSEM perspective, the model can be described as follows. Equation (43) is replaced by

$$Y = \nu + (\Lambda + \Gamma_2 X)\eta + \Gamma_1 X + \varepsilon \tag{46}$$

while Equations (44–45) remain the same, and all parameters Γ_1 and Γ_2 are given ALF or LASSO priors. The role of Γ_1 and Γ_2 are to discover measurement non-invariance across the continuum of covariates X, i.e., identifying DIF along the space of covariates. The multiple group analog of this model is the alignment method, which also allows non-invariance in the loadings and intercepts across groups, while ALF priors are given to the differences in the loadings and intercepts. There is one key difference between MNLFA and multiple group alignment (MGA). In the alignment method, the penalty function includes all pairwise differences in the parameters, while if we use the MNLFA model with dummy indicators for the groups, the penalty function includes only the differences between each group and the reference group (incomplete penalty). This can also lead to MNLFA dependence on the reference group specification.

Note also that if we remove the covariate effect on the loadings Γ_2 (metric invariance), we essentially obtain the MIMIC regularization model discussed in Asparouhov and Muthén (2024) Section 4.3. That model is a pure PSEM model which is not identified. If all direct effects Γ_1 are included in the model then the covariate effects on the factor B_1 are not identified parameters without a penalty. Thus adding ALF priors for Γ_1 yields a PSEM model. The PSEM model has the advantage over the regularization model that we don't need to sacrifice any part of the data fit. We do not need to consider the penalty weight and can let it be determined numerically.

The same applies to the loading and variance parameters to some extent. Consider the MNLFA regularization model without the mean effect B_1 but with Γ_1, i.e., unrestricted covariates effect on the indicators and without a penalty function. Consider also the following alternative parameterization of the MNLFA loading model

$$Y = \nu + \text{Exp}(\Lambda_0 + \Gamma_3 X)\eta + \Gamma_1 X + \varepsilon. \tag{47}$$

This model has all positive loadings but there is no loss of generality because the signs of the indicators can be reversed. The model is not an exact reparameterization of (46) but is sufficiently close. Using the approximation $\text{Exp}(a + bx) \approx \text{Exp}(a)(1 + bx)$ which is valid for small values of x, we see that as long as the loading non-invariance is relatively small, models (46) and (47) are approximately equal. Now, if B_1 is excluded from the model, then (47) is also a pure PSEM model and B_2 can not be identified without a penalty function. Thus adding ALF priors for Γ_2 yields a pure PSEM model.

Unfortunately, when both pure PSEM models discussed above are combined together (both mean and variance/loading effects), they do not produce a pure PSEM model. The loading and mean effects combine together and break the non-identification. Thus both (46) and (47) are technically identified models even without priors. However, these models are sufficiently close to a PSEM models and can essentially be treated as such. The identification of B_1 and B_2, without the penalty function for Γ_1 and Γ_2 is going to be poor and a null model with $B_1 = B_2 = 0$ can be used as the benchmark for the data fit and the determination of the penalty weight.

Next, we will illustrate both the MNLFA and the exponential MNLFA with a simulation study. The data is generated according to a MNLFA model. We use a 5-indicator, 1-factor and 1-covariate model. The model has one loading non-invariance on the first indicator and one intercept non-invariance (direct effect) on the second indicator. The simulation study is based on the external Monte Carlo facility in Mplus which generates the data in one step and analyzes it in a different step. The easiest way to generate such data in Mplus is with a two-level setup where each cluster contains just one observation and the variances for all (between level) random effects is fixed to 0. Figure 31 shows the Mplus input file for the data generation. Figures 32 and 33 show the Mplus input files for the external Monte Carlo estimation for the MNLFA and the exponential MNLFA respectively. The results of the simulation for the two models are given in Figures 34 and 35. For both models the bias is minimal and the coverage is near the nominal levels. A slightly smaller bias is seen for the MNLFA. To some extent this is expected since this model has the unfair advantage that it corresponds exactly to how the data is generated. The exponential MNLFA appears to have more power in detecting significance for the Γ_2 and B_2 coefficients. Note that the two parametrizations are different and therefore quantities such as bias and MSE are not directly comparable. The power to detect significance, however, is comparable. For the two non-zero coefficients in Γ_2 and B_2, the power to detect significance estimated with the exponential MNLFA is 68 and 67%.

```
MONTECARLO:  NAMES ARE y1-y5 x;
             NOBS = 500;
             NREP = 100;
             NCSIZES = 1;
             CSIZES = 500(1);
             BETWEEN = x;
             save=mnlfa*.dat;
             repsave=all;

MODEL POPULATION:
            %WITHIN%
            s1-s5 | f by y1-y5;
            logv | f; y1-y5*1;

            %BETWEEN%
            [logv@0]; logv on x*0.3;
            [s1-s5@1];
            s1 on x*0.2;
            y2 on x*0.3;
            s1-s5@0 y1-y5@0;
            f on x*0.3; f@0;
            logv@0;
            x*1;

ANALYSIS:   TYPE = TWOLEVEL RANDOM;
            ESTIMATOR = BAYES;
            fbiter=100;
```

Figure 31. Generating data for MNLFA.

```
variable:    NAMES ARE x y1-y5;
             constraint=x;

data: file=mnlfalist.dat; type=montecarlo;

MODEL:
       f by y1-y5*1 (L1-L5);
       f (v);
       f on x*0.3;
       y1 on x*0 (b1);
       y2 on x*0.3 (b2);
       y3-y5 on x*0 (b3-b5);
       [y1-y5*0];

model constraints:
       new(a*0.3);
       v=exp(a*x);
       new(LL1-LL5*1 c1*0.2 c2-c5*0);
       do(#,1,5) L#=LL#+c#*x;

model prior:
       b1-b5~ALF(0,1);
       c1-c5~ALF(0,1);
```

Figure 32. MNLFA.

For the MNLFA model, the corresponding values are 59 and 51%. The estimation of the MNLFA model takes approximately 1 second per replication.

In this simulation study, the average log-likelihood values for the MNLFA and the exponential MNLFA are −3984.8 and −3985.2 respectively. This confirms our expectation that in terms of data fit the two models are similar. The null model log-likelihood is −3987.4. This also confirms our expectation that the MNLFA model is not exactly a pure PSEM model and the two additional parameters in B_1 and

```
variable:    NAMES ARE x y1-y5;
             constraint=x;

data: file=mnlfalist.dat; type=montecarlo;

MODEL:
       f by y1-y5*1 (L1-L5);
       f (v);
       f on x*0.3;
       y1 on x*0 (b1);
       y2 on x*0.3 (b2);
       y3-y5 on x*0 (b3-b5);
       [y1-y5*0];

model constraints:
       new(a*0.3);
       v=exp(a*x);
       new(LL1-LL5*0 c1*0.2 c2-c5*0);
       do(#,1,5) L#=EXP(LL#+c#*x);

model prior:
       b1-b5~ALF(0,1);
       c1-c5~ALF(0,1);
```

Figure 33. Exponential MNLFA.

B_2 are indeed identified. The null model can still be used as a benchmark. The penalty weight should be low enough so that MNLFA yields log-likelihood that is at least as good as the null model log-likelihood.

We conclude that the MNLFA model has a great potential for identifying measurement non-invariance with respect to a variety of covariates, particularly continuous covariates and situations where multiple covariates are considered simultaneously. We would not recommend the MNLFA model as a replacement of the multiple group alignment method due to the MNLFA incomplete penalty function and dependence on the reference group. Note, however, that the MNLFA penalty function can be adjusted to match that of the alignment method precisely. The combination of Alignment and MNLFA in the presence of grouping and continuous covariates is also possible in this framework.

The MNLFA model can be used with more than one factor. All factor variances and covariances can be modeled as functions of the covariates. The most common approach is to use a linear predictor for the inverse hyperbolic tangent of the correlations. Consider the case of a model with 2 factors. If the two factor variances are v_1 and v_2, the covariance between the two factors is c, the correlation between the two factors is ρ, and the covariate is X, the model is given by

$$\log(v_1) = \alpha_1 + \beta_1 X \tag{48}$$

$$\log(v_2) = \alpha_2 + \beta_2 X \tag{49}$$

$$\tanh^{-1}(\rho) = 0.5\log\left(\frac{1+\rho}{1-\rho}\right) = \alpha_3 + \beta_3 X. \tag{50}$$

This results in the following variance covariance matrix

$$v_1 = \mathrm{Exp}(\alpha_1 + \beta_1 X) \tag{51}$$

```
                          ESTIMATES              S. E.     M. S. E.  95%  % Sig
              Population  Average   Std. Dev.   Average              Cover Coeff

 F        ON
  X            0.300     0.3152     0.1318     0.1337     0.0174 0.910 0.780

 Y1       ON
  X            0.000    -0.0321     0.1202     0.1231     0.0153 0.940 0.060

 Y2       ON
  X            0.300     0.2744     0.1283     0.1290     0.0169 0.930 0.780

 Y3       ON
  X            0.000    -0.0141     0.1227     0.1243     0.0151 0.940 0.060

 Y4       ON
  X            0.000    -0.0124     0.1088     0.1219     0.0119 0.950 0.050

 Y5       ON
  X            0.000    -0.0199     0.1214     0.1221     0.0150 0.950 0.050

 New/Additional Parameters
  A            0.300     0.3079     0.1767     0.1857     0.0310 0.950 0.510
  LL1          1.000     1.0045     0.0613     0.0642     0.0037 0.980 1.000
  LL2          1.000     1.0015     0.0571     0.0621     0.0032 0.970 1.000
  LL3          1.000     0.9968     0.0618     0.0623     0.0038 0.940 1.000
  LL4          1.000     0.9937     0.0670     0.0615     0.0045 0.940 1.000
  LL5          1.000     1.0054     0.0570     0.0619     0.0032 0.960 1.000
  C1           0.200     0.1921     0.0921     0.1028     0.0085 0.950 0.590
  C2           0.000    -0.0083     0.0844     0.0857     0.0071 0.970 0.030
  C3           0.000     0.0062     0.0881     0.0885     0.0077 0.950 0.050
  C4           0.000    -0.0015     0.0807     0.0891     0.0064 0.970 0.030
  C5           0.000    -0.0044     0.0862     0.0880     0.0074 0.980 0.020
```

Figure 34. MNLFA results.

```
                          ESTIMATES              S. E.     M. S. E.  95%  % Sig
              Population  Average   Std. Dev.   Average              Cover Coeff

 F        ON
  X            0.300     0.3240     0.1391     0.1331     0.0197 0.920 0.870

 Y1       ON
  X            0.000    -0.0520     0.1288     0.1283     0.0191 0.930 0.070

 Y2       ON
  X            0.300     0.2645     0.1366     0.1312     0.0197 0.930 0.820

 Y3       ON
  X            0.000    -0.0238     0.1296     0.1253     0.0172 0.950 0.050

 Y4       ON
  X            0.000    -0.0221     0.1165     0.1212     0.0139 0.960 0.040

 Y5       ON
  X            0.000    -0.0297     0.1275     0.1231     0.0170 0.950 0.050

 New/Additional Parameters
  A            0.300     0.3184     0.1468     0.1576     0.0217 0.960 0.680
  LL1          0.000    -0.0111     0.0609     0.0639     0.0038 0.970 0.030
  LL2          0.000    -0.0033     0.0559     0.0613     0.0031 0.970 0.030
  LL3          0.000    -0.0088     0.0618     0.0617     0.0039 0.940 0.060
  LL4          0.000    -0.0118     0.0688     0.0611     0.0048 0.930 0.070
  LL5          0.000     0.0002     0.0555     0.0608     0.0030 0.950 0.050
  C1           0.200     0.1715     0.0705     0.0817     0.0057 0.970 0.670
  C2           0.000    -0.0117     0.0716     0.0773     0.0052 1.000 0.000
  C3           0.000     0.0012     0.0742     0.0796     0.0055 0.970 0.030
  C4           0.000    -0.0047     0.0721     0.0783     0.0052 1.000 0.000
  C5           0.000    -0.0079     0.0678     0.0782     0.0046 0.980 0.020
```

Figure 35. Exponential MNLFA results.

$$v_2 = \text{Exp}(\alpha_2 + \beta_2 X) \tag{52}$$

$$c = \text{Exp}((\alpha_1 + \alpha_2)/2 + (\beta_1 + \beta_2)X/2)\frac{\text{Exp}(2\alpha_3 + 2\beta_3 X) - 1}{\text{Exp}(2\alpha_3 + 2\beta_3 X) + 1}. \tag{53}$$

The above equations must be specified precisely in the Mplus model constraint section for the MNLFA estimation. If the loadings in the factor model are all free and the factor variances are meant to be fixed to 1 in the absence of covariates, the parameters α_1 and α_2 must be fixed to zero (removed).

19. Conclusion

In this paper we illustrate the power of the PSEM framework when it is combined with various other modeling frameworks such as multilevel and mixtures. It is clear that adding a penalty to existing estimation methods can be used to efficiently obtain novel methodology that can not be otherwise accessed with traditional methods. The addition of the penalty does not lead to complications for the estimation and the added setup is minimal. Further exploration of this methodology is clearly necessary as well as real data applications.

As we have now expanded the PSEM application area, we have inevitably broached the limit of pure PSEM applications. RegSEM models are needed in many instances. This means that additional methodology is needed to guide the balancing act of sacrificing a portion of the data fit for the benefit of the penalty, i.e., the parsimony of the model. Additional development is needed to be able to effectively view and understand the continuum of models, see Asparouhov (2023) page 73, that a varying penalty weight provides. In the RegSEM examples we discussed, we stayed in the realm of minimal data fit drop, by minimizing the penalty but keeping the log-likelihood change to less than 1. This clearly would not be a universally acceptable strategy. Real-data applications usually reveal complex modeling problems and minor model modifications as in our simulation studies aren't the likely final outcome. Larger penalty weight and bigger data fit drops are likely to be of value in real data analysis. This raises questions such as how much can we increase the Geomin penalty weight and drop the data fit to obtain a more replicable model with fewer spurious cross loadings. New methodologies are needed to systematically consider the realm of RegSEM. Linear regression LASSO provides a guiding light in this regard. Many methods have been developed for linear LASSO, such as cross-validation techniques. These methods need to be expanded to the general latent variable modeling framework.

Ultimately, the penalty function is a quantification of our expectations for what the model should look like. We can argue that when the penalty is not used directly in the analysis, but only subconsciously in the analyst modeling iterations, it is much more likely for a subjective error to occur. As a consequence, replicability of the analysis will be in question. The PSEM framework and the regularization of latent variable models paired with meaningful cross-validation has the potential to greatly improve the quality of our analyses.

References

Asparouhov, T. (2023). *New features in Mplus*. Mplus workshop at Modern Modeling Methods conference at University of Connecticut. https://www.statmodel.com/download/M3TeachingSlides.pdf

Asparouhov, T., & Muthén, B. (2014a). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling*, *21*, 329–341. https://doi.org/10.1080/10705511.2014.915181

Asparouhov, T., & Muthén, B. (2014b). Multiple-group factor analysis alignment. *Structural Equation Modeling*, *21*, 495–508. https://doi.org/10.1080/10705511.2014.919210

Asparouhov, T., & Muthén, B. (2024). Penalized structural equation models. *Structural Equation Modeling*, *31*, 429–454. https://doi.org/10.1080/10705511.2023.2263913

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, *14*, 101–125. https://doi.org/10.1037/a0015583

Belzak, W. C. M., & Bauer, D. J. (2024). Using regularization to identify measurement bias across multiple background characteristics: A penalized expectation-maximization algorithm. *Journal of Educational and Behavioral Statistics*, *49*, 976–1012. https://doi.org/10.3102/107699862312264

Clark, S. L., Muthén, B., Kaprio, J., D'Onofrio, B. M., Viken, R., & Rose, R. J. (2013). Models and strategies for factor mixture analysis: An example concerning the structure underlying psychological disorders. *Structural Equation Modeling*, *20*, 681–703. https://doi.org/10.1080/10705511.2013.824786

Clark, S., Muthén, B., Kaprio, J., et al. (2009). *Models and strategies for factor mixture analysis: Two examples concerning the structure underlying psychological disorders.* https://www.statmodel.com/download/FMA%20Paper_v142.pdf.

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling*, *23*, 555–566. https://doi.org/10.1080/10705511.2016.1154793

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley and Sons.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis. Texts in statistical science series* (second edition). Chapman & Hall/CRC.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*, 21–39. https://doi.org/10.1037/1082-989X.10.1.21

Lubke, G. H., & Muthén, B. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling*, *14*, 26–47. https://doi.org/10.1080/10705510709336735

Lubke, G. H., Muthén, B., Moilanen, I. K., McGough, J. J., Loo, S. K., Swanson, J. M., Yang, M. H., Taanila, A., Hurtig, T., Järvelin, M.-R., & Smalley, S. L. (2007). Subtypes versus severity differences in the attention-deficit/hyperactivity disorder in the northern Finnish birth cohort. *Journal of the American Academy of Child and Adolescent Psychiatry*, *46*, 1584–1593. https://doi.org/10.1097/chi.0b013e31815750dd

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at ´ the Big Five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*, 471–491. https://doi.org/10.1037/a0019227

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335. https://doi.org/10.1037/a0026802

Muthén, L., & Muthén, B. (1998-2020). *Mplus user's guide*, *version 8*. 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. https://doi.org/10.1080/10705510701575396

Shedden, K., & Zucker, R. A. (2008). Regularized finite mixture models for probability trajectories. *Psychometrika*, *73*, 625–646. https://doi.org/10.1007/s11336-008-9077-9

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x