Penalized Structural Equation Models

Tihomir Asparouhov & Bengt Muthén

View supplementary material ☐

Published online: 19 Dec 2023.

Submit your article to this journal ☐

View related articles ☐

View Crossmark data ☐

Routledge
Taylor & Francis Group

Check for updates

Penalized Structural Equation Models

Tihomir Asparouhov and Bengt Muthén*

MPLUS

ABSTRACT

Penalized structural equation models (PSEM) is a new powerful estimation technique that can be used to tackle a variety of difficult structural estimation problems that can not be handled with previously developed methods. In this paper we describe the PSEM framework and illustrate the quality of the method with simulation studies. Maximum-likelihood and weighted least squares PSEM estimation is discussed for SEM models with continuous and categorical variables. We show that traditional EFA, multiple group alignment (MGA), and Bayesian SEM (BSEM) are examples of PSEM. The PSEM framework also extends standard SEM models with the possibility to structurally align various model parameters. Exploratory latent growth models, also referred to as Tuckerized curve models, can also be estimated in the PSEM framework and are illustrated here with simulation studies and an empirical example.

1. Introduction

In this note we illustrate how the penalized maximum likelihood (PML) method can be used to estimate new structural equation models. We call these models penalized structural equation models (PSEM). PML estimation is maximum likelihood estimation with a prior. The concept of prior is the same as in Bayesian estimation. The log-likelihood of the prior, which we also refer to as the penalty function, is added to the log-likelihood of the data to obtain the penalized log-likelihood. The penalized log-likelihood is then maximized to obtain the PML model estimates.

PML has been used for SEM estimation for example in Lüdtke et al. (2021), where the effect of the prior is used to stabilize the estimation in non-asymptotic settings. This is how PML is traditionally used in other branches of statistics, see Tibshirani (1996). PML has also been used for SEM in Jacobucci et al. (2016) and Huang, Chen and Weng (2017). Jacobucci et al. (2016) used the term Regularized SEM. In general, there is no difference between a regularized and a penalized model. Both terms are used interchangeably in statistical methodology to describe PML model estimation. Nevertheless, the PSEM framework that we describe here is quite different from regularized SEM (RegSEM). This framework is much more similar to the BSEM (Bayesian structural equation modeling) framework, Muthén and Asparouhov (2012), the ESEM (exploratory structural equation modeling) framework, Asparouhov and Muthén (2009), and the ASEM (aligned structural equation modeling) framework, Asparouhov and Muthén (2014) and Asparouhov and Muthén (2023b). In fact, it can be viewed as the natural extension and generalization of these different frameworks. The main difference in this new approach is that we estimate unidentified models, the prior variance and penalty regularization parameter to a large extent is irrelevant, and the penalty/prior is derived from the nature of the model and is not related to the data. The penalty reflects a complex modeling concept and it is often not just the sum of parameters' penalties as in lasso and ridge regressions.

The asymptotic behavior of the PML estimation has not generated a lot of interest in the past. When the sample size increases, the influence of the penalty/prior diminishes and the PML estimates become asymptotically equivalent to the ML estimates. This is typically the case in the RegSEM framework. In the PSEM framework, however, this is not the case. The ML estimates for a PSEM model are not available and are not uniquely identified. Asymptotically, the penalty does not disappear in the PSEM framework.

The PSEM models most often are based on the functional form of the alignment loss $\sqrt{|x|}$, instead of the traditionally used ridge loss functional form x^2 and lasso loss functional form $|x|$. Alignment loss rewards parsimony much more aggressively than ridge and lasso. This is essential for PSEM's ability to deliver novel consistent structural estimation. In this article, we do not pursue an extensive comparison of alignment/lasso/ridge. However, as it will become clear later on, the difference between alignment and lasso/ridge is very similar to the difference between the Geomin rotation and Quartimax rotation: Geomin is more aggressive in rewarding simplicity than Quartimax. It produces higher quality estimation for more complex models. Nevertheless, for many PSEM models, alignment and lasso loss functions yield nearly identical results.

The goal of the PSEM framework is to develop new structural modeling concepts that are previously unavailable due to technical limitations. Examples of such concepts are:

- Longitudinal alignment of SEM parameters. Here the alignment concept arises from the belief that parameters should be invariant across time, rather than across groups. The concept of measurement invariance is one such longitudinal example. However, longitudinal alignment can be used in models without latent variable measurement models, such as, for example, latent growth curve models (LGCM). The alignment in LGCM is an alignment for the means along a linear or quadratic curve with the goal of minimizing deviations from that curve.
- Structural alignment of SEM parameters. The concept of structural alignment is similar to the alignment in the ASEM framework but the justification does not arise from the multiple group comparison but from the structure of the model. For example, consider the direct effects in a MIMIC model. The MIMIC model generally postulates that the effect of the covariates on the dependent variables comes through the effect on the factors and any deviations from that should be minimal. We align the effects of the covariates on the dependent variables so that the number of significant direct effects is minimized. This process is also similar to BSEM but with PML the focus remains on the actual number of significant direct effects rather than their size.
- Structural modeling for EFA or aligned factors. Such structural models are available in the ESEM/AESEM frameworks but these are limited in scope due to the technical implementation. The PSEM framework provides unlimited structural modeling for EFA or aligned factors. For example, auto-regressive and growth modeling for such factors.
- Non-orthogonal and non-oblique rotations, such as those that appear in second order EFA and bi-factor EFA with more than one general factor.
- Scalar invariant multiple group EFA with orthogonal factors in all groups.
- Approximate invariance and approximately zero concepts developed with BSEM are now available with ML in the PSEM framework

This is not an exhaustive list. In this article we provide penalty functions to illustrate the above concepts. However, there is no limit to what penalty functions can be constructed and what methodological solutions can be achieved with the PML method.

Almost all of the examples we discuss in this article can be described as follows. An unidentified model M1 is desired. The model is identified by the addition of a substantively constructed prior/penalty function. The prior/penalty function is converted into implied complex model parameter constraints that directly remove the unidentified dimensions. Typically, the prior/penalty functions are based on the concept of model simplicity. They are not based on prior knowledge of the model parameters. They are based on the conceptual desire to find the simplest and most well fitting model. The observed data log-likelihood and data fit of the M1 model are the same as those of a well-known identified but uninterpretable M0 model. This modeling pattern can be illustrated well with the EFA model, where the prior/penalty is the rotation function, the M0 model is the uninterpretable unrotated factor model and the M1 model is the EFA model. Similarly, in multiple group alignment, the prior/penalty is the alignment loss function, the M0 model is the configural factor analysis model which is interpretable but can not be used for multiple group comparisons and the M1 model is the aligned model. The PML method allows us to repeat this estimation pattern for a variety of models.

It is interesting to note here that PML is a fairly straightforward and technically minor extension of the ML estimation, as compared to EFA and multiple group alignment. Standard errors for the EFA model were developed many years after the point estimation was developed because of the technical challenges. PML accomplishes this with little to no effort. Simply adding the rotation function as the model prior/penalty accomplishes the same result. It is this simplicity that allows the PML method to accomplish more advanced modeling tasks than the previously developed estimation techniques.

A key difference between BSEM and PSEM frameworks is the addition of alignment priors as an alternative to normal priors. Normal priors tend to minimize mean squared error and thus tend to spread around parameter misfit across the entire model. Alignment priors on the other hand tend to minimize the number of parameters that deviate from pure model alignment. This quality of the alignment priors delivers unbiased estimates that normal priors can not. Another advantage of PSEM over BSEM is that the priors in PSEM do not need to be conjugate priors. For example, normal priors can be given for covariances without affecting the efficiency of the estimation.

In the BSEM framework, the strength of the small variance priors is a key feature and it is often necessary to estimate the model multiple times with various prior variances, see Asparouhov et al. (2015). The PSEM framework is somewhat less dependent on that prior variance. The logic behind this is given below. Intuitively, the prior variance in PSEM should be big enough (the penalty should be small enough) as to not interfere with the data log-likelihood optimization (the optimized log-likelihood for M1 and M0 should be very close), but it should be small enough (penalty should be big enough) as to not cause numerical problems with the maximization of the penalty. If the penalty is too small, the optimization algorithm will need extraordinary precision and very small convergence criterion. Thus the size of the penalty is ideally within a certain range and within that range the PML estimates should be fairly independent of the prior variance.

Successful PML application requires skillful selection of the penalty function / parameter priors. For the examples described below, however, the penalty function is well-

known, i.e., ingenuity is not required. To obtain the Geomin rotation, we use the Geomin penalty function. Parameters that are expected to be zero have a prior with mean 0. Parameters that are expected to be equal get a zero mean prior for the difference between the parameters. In that respect, the prior setup is uniquely determined by the model description and can be considered routine.

The weighted least squares method, traditionally used for SEM estimation with categorical variables, can also be used to estimate PSEM models. As with PML, the penalty function is added to the least squares fit function and the total is minimized to obtain the penalized weighted least squares (PWLS) estimation. For both, PWLS and PML, asymptotic or bootstrap standard errors can be obtained as well as a chi-square test of fit.

2. Theoretical Justification

As a first step, we conceptualize the EFA and MGA estimation methods into a general framework. These estimation methods are essentially two-stage estimation methods. The two-stage nature is automated in software implementations and it may be perceived as a one-stage estimation. We explicate some of these details and use it to form a general methodological framework.

Let M1 be an unidentified model with parameters θ_1 and log-likelihood function $L(\theta_1)$. Suppose that a model reparameterization G separates the identified dimensions from the unidentified. That is, $\theta_1 = G(\theta_2, \theta_3)$ is a one-to-one reversible parameter transformation where θ_2 represents the parameters that can be identified and θ_3 represents the parameters that can not be identified. For example, in the EFA case, θ_1 is the loading matrix and the factor correlations, θ_2 is the unrotated loading matrix (with zero entries above the main diagonal) and θ_3 is the rotation matrix. The transformation is simply rotating the loading matrix and factor correlation matrix. In the multiple group alignment case, θ_1 represents indicator intercepts and loadings as well as the factor means and variance covariance, θ_2 represents the configural model parameters, and θ_3 represents the factor mean and variances. Fixing the unidentified parameters θ_3 to appropriate values can yield an identified model which we call the null model M0. For EFA, θ_3 is fixed to the identity rotation matrix, while in the alignment case the factor means are fixed to 0 and the factor variances are fixed to 1. In the EFA case, the null model M0 is the unrotated solution, while in the alignment case the M0 model is the configural model. Model M0 has some important properties. The log-likelihood and model fit for M0 and M1 are identical, including the defacto number of free parameters. For the M0 model, the θ_3 parameters are identified simply because they are fixed. For the M1 model, the θ_3 parameters are identified through a penalty/auxiliary function $P(\theta)$ such as the rotation simplicity function or the alignment loss function. The model estimates for M1 are obtained by maximizing the log-likelihood function $L(\theta)$ and then minimizing $P(\theta)$ in the subspace of model parameters where $L(\theta)$ is maximized. The main difference between the M1 model and

the M0 model is that the M1 model is structurally more interpretable. In the rotation case, it allows for more meaningful factor interpretation because of the simpler loading structure. In the alignment case, it allows latent factor comparison across groups.

The estimation of M1 is fairly challenging because of the two-stage optimization. As a first step, the log-likelihood $L(\theta)$ is maximized. Because the model is unidentified, the maximization amounts to determining the entire parameter subspace for which the log-likelihood is at its maximum. That parameter subspace can be somewhat challenging to describe or work with. For example, in the EFA case, the subspace consists of all orthogonally or oblique rotated loading matrices of the unrotated solution. Because that parameter subspace is obscure, the second stage of the optimization becomes challenging. As a result of these complexities, standard errors for the parameter estimates can be challenging as well. Here we introduce the PML estimation as a simpler alternative to the two-stage optimization method. Because the PML estimation is simpler, we are able to expand on both: the M1 model and the penalty $P(\theta)$. In fact, both of these can be arbitrary but the combination of M1 model and penalty function is always strategically chosen.

First let's consider the equations that determine the two-stage optimization parameter estimates. For this purpose it is easy to work with the (θ_2, θ_3) parameterization. Since the log-likelihood is independent of θ_3, the parameters that maximize the likelihood need to satisfy the following equation

$$\frac{\partial L(\theta_2, \theta_3)}{\partial \theta_2} = 0. \qquad (1)$$

This equation determines the solution for θ_2 uniquely and we denote this solution by $\hat{\theta}_2$. Minimizing the penalty then amounts to solving

$$\frac{\partial P(\hat{\theta}_2, \theta_3)}{\partial \theta_3} = 0 \qquad (2)$$

for θ_3. This equation determines the solution for θ_3 uniquely, provided that the penalty function is properly set (so that P has a unique minimum). We denote the solution by $\hat{\theta}_3$. We also denote by $\hat{\theta}_1$ the two-stage estimates obtained in the original M1 parameterization which are simply computed as $G(\hat{\theta}_2, \hat{\theta}_3)$. Note that the two stage optimization can be viewed also as a one stage constrained optimization, where Equation (2) is the parameter constraint.

Now we formally introduce the PML estimation method. Instead of maximizing the log-likelihood and then conditionally minimize the penalty in two separate steps, PML maximizes the weighted penalized log-likelihood

$$L_w(\theta_1) = (1 - w)L(\theta_1) - wP(\theta_1). \qquad (3)$$

Here w represents the weight of the penalty function and is often referred to as the regularization parameter. We are generally interested in weight values w that are small positive numbers, such as 0.1 and 0.01. If we denote by $\hat{\theta}_{1,w}$ the parameter vector that maximizes $L_w(\theta_1)$, we will show that

$$\lim_{w \to 0} \hat{\theta}_{1,w} = \hat{\theta}_1. \tag{4}$$

The above equation is the fundamental justification of the PML method and the PSEM framework. It simply states that we can obtain an increasingly good approximation of the two-stage estimation by choosing smaller and smaller penalty function weights w. Intuitively, if w is a very small number, numerically equivalent to 0, the maximization process will emphasize the maximization of L as the penalty function will be weighted down numerically to 0. Nevertheless, for a complete optimization, the penalty will also be minimized right in the vicinity where L is maximized because w is not zero.

The proof of Equation (4) is easiest to accomplish in the (θ_2, θ_3) parameterization. If $\hat{\theta}_{2,w}$ and $\hat{\theta}_{3,w}$ maximize L_w, the following equations must be satisfied

$$0 = \frac{\partial L_w(\theta_2, \theta_3)}{\partial \theta_2} = (1-w)\frac{\partial L(\theta_2, \theta_3)}{\partial \theta_2} - w\frac{\partial P(\theta_2, \theta_3)}{\partial \theta_2} \tag{5}$$

$$0 = \frac{\partial L_w(\theta_2, \theta_3)}{\partial \theta_3} = -w\frac{\partial P(\theta_2, \theta_3)}{\partial \theta_3} \tag{6}$$

Clearly (6) is the same as (2), while (5) becomes equivalent to (1) as w goes to 0. Therefore, as w goes to 0, $\hat{\theta}_{2,w}$ and $\hat{\theta}_{3,w}$ will converge to $\hat{\theta}_2$ and $\hat{\theta}_3$ and Equation (4) follows.

It should be noted here that there is a gap between a theoretical Equation (4) and actual numerical estimation. A maximization routine typically iterates through the parameter space until all parameter derivatives are numerically close enough to 0, i.e., are smaller than a particular number such as 0.000001. This number is often referred to as the convergence criterion. If we set $w = 10^{-16}$ for example, the weight of the penalty will be so low that numerically it will be eliminated. The convergence criterion will not be sufficiently small for penalty optimization to occur. The optimization will be numerically equivalent to simply optimizing the likelihood alone which will not be sufficient to identify all parameters. Ideally, the weight should be small enough so that the log-likelihoods for M1 and M0 are less than 1 apart, but not too small as to cause incomplete convergence, unidentified model, or saddle points in the standard error estimation, which is also a sign that the optimization is incomplete. Mplus reports the penalty ratio (prior proportion) $w P(\hat{\theta}_1)/(w P(\hat{\theta}_1) + (1-w)\chi^2)$, where χ^2 is the test of fit for the model. This ratio should ideally be between 0.01 and 0.10 to achieve numerically optimal results, where the PML estimation is a sufficiently good approximation for the two-stage estimation. Note, however, that in some situations χ^2 is 0 and the penalty ratio becomes 1 automatically. In such situations a different method is needed to determine an optimal range for w.

The standard errors for the PML estimator are computed by treating equations (5-6) as estimating equations. Therefore, the sandwich estimator is used, which corresponds to the MLR estimator in Mplus. The sandwich estimator is used for the PML standard error estimation even when the estimation option is set to ML. For PML, the difference between the ML and MLR estimators is only in the test of fit. MLR uses a chi-square correction factor as usual, while ML does not.

The theoretical justification for PSEM extends to the weighted least squares estimator with categorical variables. The PWLS estimator minimizes the penalized least squares fit function

$$(\sigma(\theta) - \hat{\sigma})W^{-1}(\sigma(\theta) - \hat{\sigma})^T + wP(\theta). \tag{7}$$

As in the WLS estimator, $\hat{\sigma}$ are the parameter estimates of the unrestricted multivariate probit model, $\sigma(\theta)$ are the corresponding quantities implied by the structural model, and W is the weight matrix for the least squares fit function. The penalty function is added with a positive sign because the fit function is minimized. Asymptotic standard errors and test of fit for PWLS are obtained as in Muthén and Satorra (1995).

3. Penalty Functions

The penalty functions in Mplus are implemented with the MODEL PRIORS statement. The penalty function associated with a corresponding prior is always computed as

$$Penalty = -log(Prior). \tag{8}$$

A model parameter can be given three different priors: normal, denoted by N, alignment loss function prior denoted by ALF, and the LASSO loss function prior denoted by LASSO.

If a parameter θ is given a normal prior $N(\mu, \nu)$, the penalty function, excluding irrelevant constant terms independent of the parameters, is given by

$$P(\theta) = \frac{(\theta - \mu)^2}{2\nu}. \tag{9}$$

This penalty function is also referred to as the ridge penalty function as in ridge regression, see Hoerl and Kennard (1970).

If θ is given an alignment prior $ALF(\mu, \nu)$, the penalty function is

$$P(\theta) = \frac{f(\theta - \mu)}{\nu}. \tag{10}$$

where

$$f(x) = \sqrt[4]{x^2 + \varepsilon} \approx \sqrt{|x|} \tag{11}$$

is the alignment function and ε is a small positive number that is used to ensure a smooth function f, i.e., continuous f'. This number is controlled by the TOLERANCE option in the ANALYSIS command in Mplus and is defaulted to 0.001. Smaller values of ε are associated with smaller point estimate bias and larger ε values with smaller standard error bias. Generally, in most cases, only the values of 0.01, 0.001 and 0.0001 are considered. If ε is not included, the penalty is no longer a smooth function and the penalized likelihood optimization can become quite complex, see Geminiani et al. (2021).

If θ is given the prior $LASSO(\mu, v)$, the penalty function is

$$P(\theta) = \frac{f(\theta - \mu)}{v}. \tag{12}$$

where

$$f(x) = \sqrt[2]{x^2 + \varepsilon} \approx |x|. \tag{13}$$

The role of ε here is the same as in the ALF prior: to ensure a smooth penalty function. If the TOLERANCE option is set to 0, the penalty function becomes identical to the traditional LASSO penalty function described in Tibshirani (1996).

Note that if the parameter θ is defined as a necessarily positive model parameter, such as the variance parameter, the ALF and the LASSO priors can be used without the TOLERANCE parameter, i.e., TOLERANCE = 0. The penalty function in that case will remain smooth in the admissible parameter space.

If multiple parameters are given priors, the total model penalty is the sum of all penalties associated with those priors. The role of the scale parameter v in the penalty function is to control the size of the penalty relative to the chi-square or log-likelihood function, i.e., it controls the weight w in (3). Smaller values of v increase the size of the penalty. Bigger values of v decrease the size of the penalty. The variance v can also be interpreted the way it is interpreted in BSEM modeling. The smaller the variance, the less likely a particular parameter will be to escape its target value.

The fundamental difference between normal, ALF, and LASSO priors is in the final model preference. Normal priors produce parameters with small deviations from their mean targets, while alignment priors produce models that are more parsimonious, i.e., as few as possible parameters that are off their mean targets, but the size of the deviation is bigger than what normal priors produce. If a model is desired that has only a few significant parameters not associated with the main structural model, then ALF priors should be used. If a model is desired that has all parameters not associated with the main structure be small in size, then normal priors should be used. The LASSO prior falls somewhere in between, although it is generally more similar to the ALF prior.

Consider as an illustration the following hypothetical example. Suppose that a model has 4 parameters: β_1, β_2, β_3 and β_4. Suppose that the log-likelihood of the model is identical for these 3 sets of parameter values: $M1(\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0.25)$, $M2(\beta_1 = \beta_2 = 0.5, \beta_3 = \beta_4 = 0)$, $M3(\beta_1 = 1, \beta_2 = \beta_3 = \beta_4 = 0)$. This can happen for example in linear regression where all 4 predictors are perfectly correlated. Suppose that we want to add a penalty (priors for the 4 parameters) to the model so that we can determine the most optimal/interpretable model. Table 1 shows the penalty values for the 3 sets of parameters using the three different types of penalty functions. Each method chooses the model with the smallest penalty. The Ridge penalty prefers M1, where all the parameters are closer to 0.

Table 1. Penalty function values.

Model	Ridge/Normal	LASSO	ALF
M1(.25, .25, .25, .25)	.25	1	2
M2(0.5, 0.5, 0, 0)	.5	1	1.41
M3(1, 0, 0, 0)	1	1	1

The ALF prior prefers M3 where only one parameter is not zero and the model can be interpreted as the simplest model. The LASSO prior in this case has equal penalty for all three models and can not be used for model identification. In this example, all 3 priors exhibit completely different patterns of preference.

Another important class of prior specification is based on the DIFF option. The DIFF option is used in Mplus with the Bayes estimator and is discussed in User's Guide example 5.33, Muthén and Muthén (2017). Prior specification $DIFF(\theta_1 - \theta_k) \sim N(0, v)$ produces a penalty function

$$P = \sum_{i=1}^{k} \sum_{j=1}^{i-1} \frac{(\theta_i - \theta_j)^2}{2v}. \tag{14}$$

We specify such priors when we want the θ_i parameters to be approximately equal. The DIFF option can also be used with ALF and LASSO priors. If prior specification $DIFF(\theta_1 - \theta_k) \sim ALF(0, v)$ or $DIFF(\theta_1 - \theta_k) \sim LASSO(0, v)$ is given, the penalty function is computed as follows

$$P = \sum_{i=1}^{k} \sum_{j=1}^{i-1} \frac{f(\theta_i - \theta_j)}{v}, \tag{15}$$

where f is either (11) or (13).

Another multivariate prior used in the PSEM framework is the Geomin prior specified for a loading matrix Λ. This prior gives a penalty function that is identical to the Geomin rotation criterion in EFA. Specifying such prior/penalty will enable us to include EFA/ESEM models in the PSEM framework. The Geomin prior for the loading matrix $\lambda_{11} - \lambda_{mp} \sim Geomin(m, v, \epsilon)$ yields a penalty function

$$P = (1/v) \sum_{i=1}^{p} \left(\prod_j \left(\lambda_{ij}^2 + \varepsilon \right) \right)^{(1/m)}. \tag{16}$$

Here the loading matrix is dimensioned p by m, the parameter v plays the role of variance and $1/v$ is the penalty weight, and ε is a small positive number defaulted to 0.01 which controls the smoothness of the penalty.

It is also possible to use an arbitrary penalty in Mplus. Suppose that $P(\theta)$ is the desired penalty. A new model parameter $q = P(\theta)$ is created using the command MODEL CONSTRAINT: NEW(q); $q = P(\theta)$;. Specifying a $LASSO(0, 1)$ prior for q then yields the desired penalty, assuming that $P(\theta)$ is positive and ε is set to 0. A general penalty approach is needed for example to specify rotation criteria that are different from Geomin.

4. Structural Alignment Examples

Most of the examples described in this section follow the logic of the BSEM analysis. Prior is specified for

unidentified parameters that are presumed to be zero but are allowed to break out of this assumption if the data fit requires it. PML structural alignment has several advantages over BSEM modeling. PML estimates based on ALF priors are less biased than BSEM estimates based on normal priors and are less dependent on the prior variance specification. The coverage for the model parameters is generally better with PML structural alignment. In addition, PML has the advantages of the ML framework which is more complete in terms of model fit evaluation. Another advantage of the PML structural alignment is that it can utilize simple univariate priors for all parameters. BSEM variance covariance modeling typically uses the multivariate Inverse-Wishart priors, which can be intricate to specify correctly, see Asparouhov et al. (2015). Another advantage is that PML can be used with models that are available only in the ML framework, such as EFA/ESEM models. Yet another advantage is that the PML estimation appears to be always quite fast, while the Bayes estimation may become slow when the priors are less restrictive. In this section we provide a general definition for structural alignment as well as several examples.

4.1. General Definition

In this section we provide a general definition for structural alignment. Three models will be considered: M_1, M_2, and M_3. M_1 will be referred to as the structural model, M_2 as the null model, and M_3 as the PSEM model. Model M_1 is nested within model M_2. The parameters for M_1 are the vectors θ_0 and θ_1, while the parameters for M_2 are θ_0 and θ_2. The parameter vector θ_0 represents the parameters that are common to both models, while θ_1 and θ_2 are model specific. Denote by p_i the size of the vector θ_i. Since M_1 is nested within M_2, the size p_2 of θ_2 is larger than the size p_1 of θ_1. The setup for structural alignment assumes that θ_1 is a structural constraint of θ_2 via any type of dimension-reducing technique. That is, model M_1 is obtained from model M_2 when θ_2 is subject to the following parameter constraints

$$\theta_2 = G(\theta_1, \theta_0), \tag{17}$$

where G is a multivariate function. In most situations, we consider the structured parameter vector θ_1 to be substantively more valuable than θ_2 and more easily interpretable.

The structural alignment model M_3 is a PSEM model which aims to preserve the data fit of the less restrictive model M_2, to preserve the structural constraint (17) as much as possible, and to estimate the substantively desirable structured parameter vector θ_1. The model parameters for M_3 are θ_0, θ_1, and θ_2'. Model M_3 consists of model M_2 combined with the following parameter constraint

$$\theta_2 = G(\theta_1, \theta_0) + \theta_2', \tag{18}$$

i.e., the new parameters θ_2' can be viewed as residuals for the structural constraint (17). The size of the vector θ_2' is the same as the size of the vector θ_2. It is easy to see that M_2 and M_3 provide the same fit to the data, i.e., the maximum likelihood value will be identical for the two models. M_3 has p_1 more parameters than M_2 and thus will have p_1

unidentified dimensions. To identify these dimensions, we use a small weight penalty function

$$P = \sum_{i=1}^{p_2} f(\theta_{2i}'), \tag{19}$$

where f is the ALF prior function (11) and θ_{2i}' is the i−th entry in the vector θ_2'. This penalty function is constructed by specifying ALF priors for all residual parameters θ_2'. The PSEM model M_3 preserves the data fit of M_2 and estimates the structural parameters θ_1 by minimizing the error in the structural constraint (17) which defines M_1.

We generally refer to model M_1 as the structural model, M_2 as the null model, and M_3 as the PSEM model. The above general definition can be applied to any structural parameter constraint formulated as in (17). In most cases, we provide only a description of the PSEM model M_3. The structural model M_1 can be obtained from M_3 by fixing all parameters with ALF priors (these are the parameters θ_2') to 0. The null model M_2 can typically be obtained from M_3 by fixing θ_1 to 0. However, identifying θ_1 may not be as obvious in all cases. Typically, these are the minimal set of M_1 parameters that must be fixed to zero to make the unidentified model M_3 into an identified model M_2. In most cases we describe below, $G(0, \theta_0) = 0$, which means that when θ_1 is fixed to 0, $\theta_2 = \theta_2'$. This can be used to determine the parameters θ_2.

In most applications, it is desirable to determine and estimate all three models. For example, by comparing the log-likelihood values of M_2 and M_3, we can ensure that the penalty weight is sufficiently small as to not affect the data fit in a substantial way. The log-likelihood values of M_2 and M_1 can be used to conduct a likelihood ratio test to determine if the parameter constraint (17) is satisfied. Alternatively, (17) can be tested within the M_2 estimation using Wald's test, i.e., MODEL TEST in Mplus. If the structural constraint (17) is not rejected, then there is no need for model M_3. If the structural constraint is rejected, however, model M_1 is rejected. Nevertheless, we can use model M_3 instead of model M_1 for inference regarding θ_1 parameters.

4.2. Latent Growth Curve Alignment of Means

Latent growth curve models (LGCM) imply a very restrictive mean structure. If a variable Y_t, $t = 1, ..., T$, is modeled for example with a linear growth curve model, T mean parameters are fitted with just two parameters: the means of the random intercept and slope. Mean parameters typically have a lot of power to reject the model when they are misfitted and that can cast doubt on the entire model even if the model is mostly accurate.

The linear LGCM is given by the following equations

$$Y_t = I + S \cdot t + \varepsilon_t \tag{20}$$

$$I \sim N(\alpha, v_{11}), S \sim N(\beta, v_{22}), Cov(I, S) = v_{12}, \varepsilon_t \sim N(0, \sigma_t). \tag{21}$$

The PSEM based LGCM model replaces Equation (20) with

$$Y_t = \nu_t + I + S \cdot t + \varepsilon_t \tag{22}$$

```
MONTECARLO:
        NAMES = Y1-Y8;
        NOBSERVATIONS = 500;
        NREPS = 100;

MODEL POPULATION:
i s | Y1@0 Y2@0.1 Y3@0.2 Y4@0.4 Y5@0.6 Y6@1 Y7@1.2 Y8@1.5;
[i*0.4 s*0.1];
i*0.9 s*0.4; i with s*0.3;
[Y1-Y4*0 Y5*-0.3 Y6*0.2 Y7-Y8*0];
Y1-Y8*1;

MODEL:
i s | Y1@0 Y2@0.1 Y3@0.2 Y4@0.4 Y5@0.6 Y6@1 Y7@1.2 Y8@1.5;
[i*0.4 s*0.1];
i*0.9 s*0.4; i with s*0.3;
[Y1-Y4*0 Y5*-0.3 Y6*0.2 Y7-Y8*0] (m1-m8);
Y1-Y8*1;

MODEL PRIORS:  m1-m8~ALF(0,1);
```

Figure 1. Simulation study for LGCM means alignment.

$$\nu_t \sim ALF(0,1) \qquad (23)$$

while (21) remains the same. We can estimate the above model when we add the zero-mean alignment prior for the time specific intercepts ν_t. The PML estimation in this case aligns the time specific means along the estimated growth curve model. This alignment resembles the MGA methodology where now the time variable plays the role of the multiple group variable. Standard SEM methodology can not estimate time specific intercepts ν_t as well as the intercepts α and β for the latent growth factors. This is because the mean sufficient statistic has T degrees of freedom (the sample means of Y_t) which can not identify the $T + 2$ parameters ν_t, α and β. In the PSEM framework, however, this is possible. The main advantage of the PSEM-LGCM is that it can estimate the growth curve model while at the same time it can fit the observed means perfectly due to the time-specific intercept parameters.

Next we show how this example fits in the general framework described in the previous section. The structural model M_1 is given by (20-21). The structural parameters θ_1 are α and β, while θ_0 are all other M_1 parameters. The PSEM model M_3 is given by (21-23) and ν_t in (22) represents θ_2'. The null model M_2 is given by (21-22) with $\alpha = \beta = 0$ and ν_t represents the parameters θ_2 in that model. The parameter constraints (17) that we provide a structural alignment for in this case is given by $\nu_t = \alpha + t\beta$, i.e., we structurally align the mean parameters ν_t along the best straight line across time.

We illustrate the performance of the PSEM methodology with a simulation study. Figure 1 shows the input file for a linear LGCM model with $T = 8$ unevenly spaced time points and $N = 500$. We generate the data so that at time points 5 and 6 the means of the observed variables deviate from the linear growth projection, i.e., $\nu_5 = -0.3$ and $\nu_6 = 0.2$. The results of the simulation study for a selection of the parameters are presented in Table 2. The bias in the parameter estimates is minimal and the coverage is near the nominal level of 95%. The PSEM methodology is able to recognize that the means at time point 5 and 6 do not conform with the linear growth model and the deviations from

Table 2. Simulation study results for LGCM means alignment.

Parameter	True Value	Abs. Bias	Coverage
ν_4	0	.01	.96
ν_5	−.3	.00	.96
ν_6	.2	.02	.93
α	.4	.00	.93
β	.1	.01	.96

the linear growth model are estimated correctly. The latent variable means are estimated correctly as well despite the fact that all Y_t intercepts are also free parameters. The average chi-square value for this PSEM model is 26 and with 25 degrees of freedom the model is rejected 8% of the time. For reference, the standard LGCM for these data yields an average chi-square of 92 and with 31 degrees of freedom this yields 100% rejection. Furthermore, in the standard LGCM, the intercept parameters α and β are biased and have lower coverage. PSEM in Mplus does not identify the null model as that is not necessary for the PSEM estimation. The degrees of freedom are computed not by identifying the null model but are computed as in Zou et al. (2007).

Note that we did not perform an investigation on selecting the weight for the penalty which manifests here in the second parameter of the ALF prior. We refer to that parameter as the variance parameter in line with the normal prior but this is not the variance of the ALF distribution. Using the prior variance value of 1 is a good starting point in most situations. Changing the variance to 10 does not affect the estimation, while setting it to 0.1 resulted in some small increase in the bias in the parameter estimates and the average chi-square value increased by 0.5. This indicates that the size of the penalty is a bit too large when the variance is set to 0.1 and that is preventing the data log-likelihood to be optimized completely. The ratio of the penalty as a proportion of the total optimization function (chi-square plus penalty) is 0.7 when the prior variance is 0.1. That quantity is 0.2 when the prior variance is set to 1, and it is 0.02 when the prior variance is set to 10. Generally, as the theoretical results indicate, we prefer the penalty to be as low as possible, and the prior variance to be as high as possible. However, it is not necessary to search for the largest variance value that will allow correct PML estimation, since the results are not sensitive to that parameter. In this example variance of 1 works well also with larger and smaller sample sizes and various metrics of the dependent variables. The most clear indicator that the variance value of 1 is sufficiently small is the fact that the PSEM model produces the same chi-square value as the null model. In practical settings, it might be helpful to estimate the null model as a preliminary step. We can then assert that the PSEM-LGCM penalty weight is OK, i.e., is sufficiently small, if the PSEM-LGCM and the null model yield chi-square values that are nearly identical.

In the above example, using normal priors instead of alignment priors produces larger biases in the parameter estimate. This emphasizes the benefit of using alignment priors and the advantages of PSEM over the corresponding BSEM Bayesian estimation.

4.3. Alignment of Covariate Effects in MIMIC Models

Let Y represent a vector of dependent variables, η represent a vector of latent variables measured by the dependent variables Y and X represent a vector of covariates. The MIMIC model, see Bollen (1989), in vector form can be expressed as follows

$$Y = \nu + \Lambda\eta + \varepsilon \qquad (24)$$
$$\eta = \Gamma X + \zeta. \qquad (25)$$

The model postulates that the effect of the covariates on the dependent variables can be explained entirely by the effect of the covariates on the factors. This assumption is often too restrictive and the model can be rejected due to the presence of direct effects from the covariates to the dependent variables. Typically, modification indices are used to discover needed direct effects, which are then added one at a time. When the model is complex, however, with a substantial number of variables, the process of sequentially adjusting the model can be time-consuming. This process can also become subjective because at each step multiple direct effects can have approximately equal modification indices. As a result of that and the imperfect nature of the modification indices, the most parsimonious model may be omitted. The PSEM framework offers a one-step solution which essentially aligns the effects of the covariates along the factor measurement model, i.e., the effect of the covariates on the indicators are aligned to be proportional to the factor loadings. The PSEM model replaces Equation (24) with

$$Y = \nu + \Lambda\eta + BX + \varepsilon \qquad (26)$$
$$B \sim ALF(0, \nu). \qquad (27)$$

Without (27), the model can not be identified. This model can also be formulated for categorical variables where the underlying Y^* variable takes the role of Y.

We illustrate the performance of the estimation for the above model with a simulation study using 10 categorical indicator variables with 4 categories each, 2 factors and 3 covariates. We use the penalized WLSMV (PWLSMV) estimation method in this simulation study. The Mplus input file for the data generation and model estimation is given in the online Supplemental Materials (SM) Figure S1. Two non-zero direct effects are included in this simulation study. The results of the simulation study are presented in Table 3 for all non-zero regression parameters. The bias in the parameter estimates is minimal and the coverage is near the nominal level of 95%. The PWLSMV estimator automatically discovers the non-zero direct effects among all possible such effects and provides unbiased estimates for these effects. All other direct effects are estimated to values that are approximately zero. The average chi-square value is 34 and with 34 degrees of freedom this results in a 3% rejection rate. If the direct effects are not included in the MIMIC model the model is rejected 100% of the time and biases appear in both Γ and Λ. In this estimation we used a prior variance of $\nu = 10$. This is necessary to reduce the penalty proportion. With prior variance of 1 the penalty proportion is 0.38 and with prior variance of 10 it is 0.06,

Table 3. Alignment results of direct effects in MIMIC models.

Parameter	True Value	Abs. Bias	Coverage
γ_{11}	0.5	.01	.95
γ_{12}	0.6	.00	1.0
γ_{13}	0.7	.02	.97
γ_{21}	0.7	.00	.95
γ_{22}	0.6	.00	.91
γ_{23}	0.5	.00	.95
β_{13}	0.3	.03	.97
β_{21}	0.4	.01	.93

which is preferable. However, the difference in the results between these two settings is negligible.

4.4. Alignment of Cross-Loadings in Confirmatory Factor Analysis: Improving Target Rotation

Confirmatory factor analysis (CFA) estimates a prespecified loading structure. The PSEM methodology can extend the CFA model by including all loadings, i.e., by including all possible cross-loadings. The additional loadings are aligned with a zero-mean alignment prior. This modeling approach is essentially an alternative to the EFA target rotation. Specifying targets in the target rotation becomes equivalent to specifying priors for these loadings in PSEM. In fact, if we use normal prior instead of alignment prior, the penalty function for PSEM becomes identical to the target rotation criterion. Therefore, subject to (4), the PSEM model with normal prior is expected to produce approximately the same results as the target EFA rotation. There are two advantages of the PSEM approach. First, we can vary the strength of the targets in PSEM. Certain targets can be assigned medium strength while other targets can be assigned full strength. This can be accomplished by using two different variance parameters. Full strength targets can be given priors with smaller variance, while medium strength targets can be given priors with larger variance. Such an approach can be helpful when there is ambiguity in the measurement model. Certain indicators may be hypothesized to be potentially related to a factor even if they are not intended to be a part of the measurement model for that factor. The second advantage of the PSEM model lies in the use of the alignment prior. The target rotation and the PSEM model with normal prior are likely to distribute error in the presumed model while the alignment based PSEM will concentrate the errors in as few zero-target cross-loadings as possible.

We illustrate the performance of the PSEM methodology with a simulation study. Data is generated with a 6 indicator 2 factor analysis model where each factor has 3 main indicators and there is just one small but non-zero cross-loading. We analyze the data using 5 methods: PSEM-ALF with alignment prior, PSEM-N with normal priors, PSEM-LASSO with LASSO priors, BSEM with normal prior, ESEM with target rotation. The targets and priors are assigned to all cross-loadings. All of these methods are designed to estimate both the main and the cross loadings. Mplus input file for this simulation study using the PSEM-ALF estimation method is given in SM Figure 2. To estimate PSEM-N and PSEM-LASSO we simply replace the ALF prior specification

with N and LASSO. To estimate the BSEM model we change the estimator to Bayes and use Normal priors. The input for ESEM-target estimation is given in SM Figure 3.

Table 4 show the results of this simulation study for the 5 estimation methods for a selection of the parameters. PSEM-ALF estimation shows no bias and coverage near the nominal level. PSEM-LASSO results are quite similar to PSEM-ALF. The differences between the two methods become more noticeable with larger sample sizes. For example, for $N = 10000$, the factor covariance estimate for PSEM-LASSO is biased and estimated at 0.29, while the PSEM-ALF estimate is unbiased at 0.25. The biases in PSEM-LASSO at that sample size also lead to drop in coverage down to 70% while the coverage for PSEM-ALF remains near the nominal levels.

The other 3 estimation alternatives show biased results and poor coverage, not just for the cross-loadings but also for the factor covariance. Most importantly, the PSEM-ALF method appears to be vastly superior to the EFA target rotation in terms of properly estimating the cross-loadings with targets. It should be noted that for this estimation problem, the ESEM-target rotation has previously been considered to be the best estimation method. Clearly the PSEM framework now offers a better alternative. The results also show that ESEM-target and PSEM-N yield almost identical results. This is an important confirmation of the theoretical result given in Equation (4). Ultimately, this also shows that EFA estimation is simply a standard ML estimation, where the rotation criterion is a penalty function or the prior for the model parameters. Later on, we will take advantage of that fact to estimate certain EFA models that are not available within the standard ESEM framework.

The BSEM method in this example performed the worst. This is in part due to the fact that we did not vary the prior variance for that estimation and that is an essential part of the proper BSEM application. With a very careful BSEM analysis, the BSEM results are expected to be as good as PSEM-N and ESEM-target, i.e., not nearly as good as PSEM-ALF.

The target rotation can also be estimated without specifying the small non-zero cross-loading as a target. In that case the target rotation performs well and yields results similar to PSEM-ALF. We conclude that the target rotation EFA is biased only when the targets are specified incorrectly. PSEM-ALF has the advantage that it gives unbiased estimates even when the targets are specified incorrectly. Using target EFA, instead of PSEM-ALF, thus carries an additional risk of specifying the targets incorrectly. It also means that, for best results, EFA target rotation should be adjusted and possibly re-estimated until all targets are not significant. If a target rotation EFA is estimated and some of the targets have statistically significant estimates, the model should be re-estimated with different targets to avoid the biases illustrated in this simulation study.

In principle, for this example, the model can be estimated with just 2 targets, one for each factor. This applies to both the ESEM-target and the PSEM-ALF estimations. In general, ESEM-target and PSEM-ALF can be estimated with as few

Table 4. Alignment of cross-loadings: Absolute bias(coverage).

Parameter	True Value	PSEM ALF	PSEM LASSO	PSEM Normal	ESEM Target	BSEM
λ_{11}	1	.00(.93)	.00(.93)	.01(.93)	.00(.93)	.02(.98)
λ_{12}	0	.00(1.0)	.00(1.0)	.01(.94)	.01(.95)	.30(.95)
λ_{61}	.3	.02(.99)	.03(.92)	.09(.27)	.09(.27)	.15(.97)
λ_{62}	1	.00(.95)	.00(.95)	.02(.91)	.02(.91)	.08(.99)
ψ_{12}	.25	.00(.96)	.02(.94)	.06(.71)	.07(.63)	.35(.98)

as $m - 1$ targets per factor where m is the number of factors in the model. Having more targets than the minimum number, however, is useful in reducing the mean squared error (MSE) of the estimates. This conclusion applies again to both the ESEM-target and the PSEM-ALF estimations. In fact, if targets are given for all zero cross-loadings, the MSE for ESEM-target and PSEM-ALF is nearly as good for this example as the corresponding CFA model. Furthermore, PSEM-ALF achieves such a level of MSE, even if alignment priors are given for all cross-loadings (even those that are not zero). This advantage does not apply to ESEM-target.

In conclusion, PSEM-ALF can be used to estimate all main loadings as well as all cross-loadings similarly to the EFA-target method. Generally, PSEM-ALF is a one-step estimation and there are no implications when a zero target turns out to be a non-zero significant cross-loading. On the other hand, for optimal performance, ESEM-target may require multiple estimations. First, ESEM-target should be estimated with the minimum number of targets. Subsequently, targets can be added for all insignificant cross-loadings. If a targeted loading becomes significant or if it is estimated to a larger non-zero value, the target for that loading should be removed. If necessary, to maintain the minimum number of targets, a target that is removed should be replaced with a different target. With every multi-stage estimation procedure, there is a risk of omitting the most parsimonious model. Since PSEM-ALF automates this process, we conclude that PSEM-ALF should be preferred over the EFA target rotation.

4.5. Alignment of Residual Correlations in EFA

Next we turn our attention to the residual correlations in EFA models. The ESEM framework can estimate such correlations but it requires preliminary analysis to determine which residual correlations should be included in the model, see Ferrando et al. (2022). It is not possible in the ESEM framework to estimate all residual correlations as the model becomes unidentified. BSEM can be used to estimate all residual correlations in CFA but not in EFA. Bayes estimation for ESEM models is currently not available. Only standard EFA models without residual correlations can currently be estimated with the Bayes estimator, see Asparouhov and Muthén (2012). The PSEM framework can offer a unique alternative: estimate the EFA model and all residual covariances at the same time. The residual covariances are given zero-mean alignment priors to identify the model. To be clear, the PSEM framework offers two variations of such estimation. It is possible to estimate the model with a penalty function which is the sum of the rotation criterion and

Table 5. Alignment results for residual correlations in EFA.

Parameter	True Value	Abs. Bias	Coverage
λ_{11}	1	.01	.94
λ_{12}	0	.02	1.0
λ_{31}	1	.02	.91
λ_{32}	.5	.00	.97
ψ_{12}	0.25	.00	.98
θ_{12}	0.5	.05	.87
θ_{13}	0	.01	1.0
θ_{48}	0.3	.02	.94
$\theta_{9,10}$	0.4	.03	.96

Table 6. Alignment results for residual correlations and cross-loadings in CFA.

Parameter	True Value	Abs. Bias	Coverage
λ_{11}	1	.00	.97
λ_{12}	0	.01	1.0
λ_{31}	1	.01	.95
λ_{32}	.5	.01	.96
ψ_{12}	0.25	.00	.98
θ_{12}	0.5	.03	.90
θ_{13}	0	.01	1.0
θ_{48}	0.3	.02	.96
$\theta_{9,10}$	0.4	.03	.94

the zero-mean alignment priors for all residual covariances. This approach is discussed in the next section. Here we illustrate a simpler alternative. Using the ESEM framework, we estimate the EFA model with a penalty function which is the sum of the zero-mean alignment prior for all residual covariances. The penalty function contains only the priors of the residual covariances, while the rotation criterion is optimized after the penalized log-likelihood maximization as in the traditional EFA estimation. This approach is simpler to specify in Mplus because the rotation criterion is not included in the penalty specification and is automated by the EFA estimation.

We illustrate this approach for an EFA model with 10 indicators and 2 factors. Each factor has 5 main measurements. One of the factors has a mid-size cross loading as well. Additionally, we generate the data with 3 non-zero residual correlations: θ_{12}, θ_{48} and $\theta_{9,10}$. The model estimation includes all residual correlations in an attempt to discover those that are not zero. The Mplus input file for this simulation study is given in SM Figure 4. The PSEM model is specified as an ESEM model, i.e., EFA model for the two factors. All residual covariances are estimated as free parameters with ALF(0,1) priors.

The results of this simulation study for some of the parameters are given in Table 5. First, we see that the EFA model is largely unaffected by the inclusion of all residual covariances. The loading structure is estimated and rotated as if the residual covariances are not included. All loading parameters, including the cross-loadings, appear to have minimal bias and coverage near the nominal level. In addition, all residual covariances, the non-zero and the zero covariances, are estimated well with small bias and good coverage. The 42 zero covariances are significant in less than 3% or the replications, while the 3 non-zero covariances are significant more than 95% of the time. We conclude that the PSEM estimation method can be used to estimate

EFA models and all free residual covariances at the same time.

Note that the null model here is the unrestricted variance covariance model. Thus, the PSEM model, EFA with all free aligned residual covariances, is expected to have an average chi-square value near 0 and it can not be used to test the EFA structure. In the above simulation study, the average chi-square value is 0.1. The model has 0 degrees of freedom. The chi-square is not precisely 0, due to the approximation depicted in Equation (4).

4.6. Simultaneous Alignment of Residual Correlations and Cross-Loadings in CFA

In this section we illustrate how PSEM simultaneously aligns the residual correlations and the cross-loadings in a CFA model. We use the same simulation setup as in the previous section, i.e., a two-factor analysis model with 10 indicators where each factor has 5 main indicators. There is one cross-loading and 3 residual correlations in the data generation and we want the PSEM estimation to automatically discover those and provide unbiased estimates and standard errors. PSEM estimates, in addition to the CFA model, all cross-loading parameters and residual covariance parameters with ALF(0,1) prior.

Note that in this model, each non-zero cross-loading is equivalent to adding 5 non-zero residual correlations. Alignment optimizes in terms of the most parsimonious model. A loading parameter should have an advantage escaping its prior mean specification over the equivalent 5 residual correlations, i.e., the penalty for having one non-zero loading should be smaller than the penalty for having 5 non-zero residual covariances.

The PSEM estimation is conceptually similar to the BSEM methodology, which can also be used for such estimation. An advantage of the PSEM method is that it requires only a single estimation. The prior variance does not need to be varied. In addition, the model converges fast as the PSEM model is never under-identified. The BSEM model is often under-identified when the prior variance is less restrictive. PSEM also produces smaller biases in the parameter estimates. BSEM tends to average out model misspecifications which results in more biased estimates.

The Mplus input file for this simulation study is given in SM Figure 5 and Table 6 shows the results for some of the model parameters. The bias for all parameters is minimal and the coverage is near the nominal level. The joint alignment of cross-loadings and residual correlations appears to work well in the PSEM framework.

Note that this PSEM model can also be viewed as a version of the EFA model with all residual covariances discussed in the previous section. Here we have combined the ALF based target rotation (alignment of cross-loadings) with the residual covariances alignment, i.e., the penalty function includes the rotation penalty as well as the residual covariance penalty.

5. Multiple Group and Longitudinal Alignment Examples

In the previous section we showed that the target rotation is equivalent to the PSEM model with normal priors and that ALF priors provide an even better version of the target rotation. In this section we similarly illustrate that the multiple group alignment (MGA) methodology is equivalent to the PSEM model with DIFF priors. This allows us to further customize MGA to various structural models or to customize the actual alignment optimization. The logic behind this is the fact that the MGA alignment loss function is identical to the penalty function given by the DIFF priors. Huang (2018), Tutz and Schauberger (2015), and Magis et al. (2015) also describe a penalized method that is akin to multiple group alignment.

As implemented in Mplus, the MGA alignment loss function weighs the loss due to difference between two parameters with weights inversely proportional to the square root of the group sizes. The purpose of this additional weighting is meant to minimize MSE of the estimates. If the groups are of equal size, the additional weighting can be ignored because all weights are equal. With unequal group sizes, the weighting is likely to provide only a marginal difference, i.e., in most practical situations this additional weighting can be ignored.

It is possible to use DIFF priors in PSEM with the MGA weights as well by specifying the corresponding variance parameter in the ALF priors. However, this becomes impractical for larger examples due to the large number of DIFF priors that will need to be specified separately. To simplify the PSEM specification for the MGA model, we will be using ALF priors for each difference with the same mean of 0 and the same variance, i.e., the additional MGA weighting will be ignored. In our illustration below, we use groups of equal sizes, which allows us to match precisely the MGA loss function and PSEM penalty derived from the DIFF priors with constant variance.

5.1. Multiple Group Alignment with PSEM

In this section we compare the MGA and the PSEM based MGA model estimations. The MGA model is given by the following equations

$$Y_{ig} = \nu_{ig} + \lambda_{ig}\eta_g + \varepsilon_{ig} \tag{28}$$

$$\eta_g \sim N(\alpha_g, \psi_g), \varepsilon_{ig} \sim N(0, \theta_{ig}). \tag{29}$$

For identification purposes $\alpha_1 = 0$ and $\psi_1 = 1$, i.e., we use a fixed alignment with the first group as a reference group. We illustrate the equivalence of the two estimators with a simulation study using a three-group factor analysis model with one factor measured by 4 variables. Data is generated with one non-invariant intercept and one non-invariant loading. The input file for the MGA simulation study is given in SM Figure 6 and for the PSEM-MGA simulation study in SM Figure 7. In the PSEM-MGA model, the prior specification is the alignment version of holding the intercept and loading parameters equal across groups. We specify ALF DIFF priors for each intercept and loading parameter.

Table 7. Multiple group alignment: Absolute bias(coverage).

Parameter	Value	Alignment	PSEM Alignment
λ_{11}	1	.01(.92)	.01(.93)
λ_{41}	1	.01(.95)	.01(.94)
ν_{11}	0	.00(.96)	.00(.95)
ν_{41}	.5	.00(.98)	.00(.98)
λ_{12}	1	.01(.95)	.01(.94)
λ_{42}	.5	.01(.89)	.00(.88)
ν_{12}	0	.01(.96)	.01(.96)
ν_{42}	0	.00(.95)	.00(.95)
α_2	.4	.00(.91)	.00(.91)
ψ_2	1.5	.02(.93)	.03(.92)

The results of the simulation study for some of the parameters are given in Table 7. Parameter estimates and standard errors are nearly identical for the two estimations. The log-likelihood and chi-square test of fit are nearly identical as well.

We conclude that MGA and PSEM-MGA methodologies are equivalent. This fact can be used to extend the capabilities of the alignment methodology. Consider for example the case where there are multiple factors in the model and the most optimal reference group for the different factors is different. Such a model can easily be accommodated in the PSEM-MGA framework while it is not available in the MGA framework.

Another example which needs customized alignment is the alignment of a factor with only two indicators. The configural model in that case is unidentified and thus the MGA approach is not possible. PSEM-MGA approach, however, can be implemented using a customized model. For example, both loading parameters can be fixed to 1 in all groups, the factor variance can be estimated as free in every group, the factor mean can be fixed to 0 in the reference group and free in all other groups, while the indicator intercepts are aligned as usual using the DIFF ALF priors. In this case, we would align only the factor means while the factor loadings are held fixed and all equal to 1.

The alignment setup with PSEM can also be extended to any structural model. Structural modeling in the ASEM framework is limited to structures that can be aligned. For example, a factor predictor in ASEM must have a free and unequal regression coefficient in every group. In the PSEM-MGA framework there is no such restriction and a group invariant regression coefficient can be estimated. Restricting the regression coefficient to be group invariant is essential if the predictor is a group level variable. Another situation that would be available only with PSEM-MGA is the possibility to use alignment only for some indicators. Indicators that are not aligned can have either full measurement invariance (intercepts and loadings can be constrained to be equal across groups) or full measurement non-invariance (intercepts and loadings are unconstrained and unequal across groups). The next section describes yet another example where the PSEM-MGA modeling can be utilized.

5.2. Metric Scale Alignment

The MGA method aligns all intercepts and loadings at the same time. The traditional path for establishing measurement

Table 8. Metric scale alignment.

Parameter	True Value	Abs. Bias	Coverage
λ_{11}	1	.01	.93
λ_{41}	1	.01	.94
λ_{12}	1	.01	.94
λ_{42}	.5	.00	.88
ψ_2	1.5	.03	.92
λ_{13}	1	.01	.92
λ_{43}	1	.01	.93
ψ_3	1.2	.03	.96

invariance, however, estimates the following sequence of models: configural, metric invariance where only the loadings are invariant and finally scalar invariance where both the loadings and the intercepts are invariant. Often scalar invariance is difficult to establish while metric invariance is not. In such situations, full alignment may have too many non-invariant intercepts and the model interpretation may become difficult. Alignment of the intercepts may also affect the alignment of the loadings since the two processes are intertwined, which would prevent us from evaluating loading invariance independently of intercept invariance. Therefore in certain situations it might be desirable to align only the loading parameters as a preliminary step before full alignment is attempted. If we only align the loadings, the factor intercepts in all groups will be fixed to 0 and the intercepts for all indicators will be estimated as free, unequal, and unaligned parameters. This model essentially aims to provide an approximation to the metric scale multiple group measurement model and we will refer to it as the metric scale alignment. This model can easily be estimated in the PSEM framework and it is not available with the traditional alignment methodology.

Using this new model it is possible to conduct a comprehensive measurement invariance analysis that is based on alignment. As a first step, metric alignment is evaluated using the metric scale alignment. Using these results, we can determine which loadings are not invariant and which are invariant. At that point, we can estimate a model where all invariant loadings are held equal to each other, all non-invariant loadings are estimated as free and unequal while all intercepts are aligned. In some situations where measurement invariance is difficult, such comprehensive measurement invariance alignment might be preferable over a simultaneous alignment method. Ultimately, this new approach must be systematically evaluated with empirical studies.

In this section we use a simulation study to illustrate the metric scale alignment for the scenario where scalar invariance does not hold at all but metric invariance holds with minor exceptions. The Mplus input file for this simulation study is given in SM Figure 8 and Table 8 shows the results. We use a 3-group analysis where 1 factor is measured by 4 variables. The intercepts of the 4 variables are different across groups and do not reflect a scalar invariant model. The factor loadings, however, are invariant with the exception of one loading parameter. The factor intercepts in the estimated model are all fixed to 0, while the factor variance is estimated as a free parameter except for the reference group. The indicator intercepts are free and unequal

parameters. The intercepts are not aligned. Only the factor loadings are aligned with the ALF DIFF priors.

The results indicate that the model parameters are recovered well. The bias for the factor variances and the loading parameters are minimal and the coverage is near the nominal level. If we analyze the same data with the full scalar alignment, almost all intercept parameters become non-invariant. The loadings and factor variance parameters, however, are estimated correctly. Thus, at least for this example the advantage of the metric alignment is simply in the easier model interpretation.

5.3. Longitudinal Alignment

Longitudinal alignment is analogous to multiple group alignment where the grouping variable is the time. When latent variables are measured repeatedly across time we can align the measurement model across time. If the latent variables are uncorrelated across time, the model can be estimated with multiple group alignment via the usual wide to long data and model transformations. The latent variables, however, are almost certainly correlated across time and thus the model must be estimated and aligned as a longitudinal model.

In this section we describe the general longitudinal alignment model and provide a simulation study to illustrate how such a model is estimated in Mplus. This modeling technique uses the MGA alignment loss function, i.e., this is not a PSEM example. In the following section we will show how the PSEM methodology can be used to estimate a growth model for the longitudinally aligned factors, i.e., PSEM provides valuable structural extensions for the longitudinal alignment model as well.

Suppose that Y_{ti} is an observed measurement for a latent factor f_t at time t, where $i = 1, ..., P$ and $t = 1, ..., T$, i.e., the latent factor is measured by P variables across T time periods. If measurement invariance holds across time we can study the changes in the factor means and variances across time. Measurement invariance, however, will often fail in practical applications and thus the alignment of the measurement model is needed. The model can be described as follows

$$Y_{it} = \nu_{it} + \lambda_{it}f_t + \varepsilon_{it} \qquad (30)$$

$$\varepsilon_{it} \sim N(0, \theta_{it}) \qquad (31)$$

$$f = (f_1, f_2, ..., f_T) \sim N(\alpha, \Sigma). \qquad (32)$$

The intercepts ν_{it} and loadings λ_{it} are free and unequal parameters that are aligned via the alignment loss function as in MGA. The vector α contains the factor means across time and the factor variance covariance matrix is an unstructured/unconstrained variance covariance matrix Σ. We denote the factor mean at time t by α_t and the covariance between f_i and f_j by σ_{ij}. The fixed alignment estimation sets for identification purposes: $\alpha_1 = 0$ and $\sigma_{11} = 1$, while the free alignment estimation sets $\sigma_{11} = 1$ while α_1 is estimated as a free parameter just as the factor means at the other time points. The longitudinal alignment extends also to structural models along the lines of the ASEM framework

Table 9. Longitudinal alignment.

Parameter	True Value	Abs. Bias	Coverage
λ_{11}	1	.01	.98
λ_{21}	1	.01	.94
λ_{12}	1	.00	.99
λ_{22}	.5	.01	.92
ν_{41}	0	.02	.96
ν_{51}	0	.00	.99
ν_{42}	.5	.02	.99
ν_{52}	0	.00	.98
α_2	.3	.00	.99
α_3	.6	.01	.98
σ_{22}	.9	.01	.95
σ_{33}	1.4	.02	.98
σ_{12}	.5	.00	.97

Table 10. Longitudinal alignment growth model.

Parameter	True Value	Abs. Bias	Coverage
λ_{11}	1	.01	.98
λ_{21}	1	.01	.96
λ_{12}	1	.01	.98
λ_{22}	.5	.00	.99
ν_{11}	0	.01	.95
ν_{21}	0	.00	.95
ν_{14}	0	.02	.96
ν_{24}	.5	.02	.95
m	.3	.01	.95
v	.3	.01	.99
c	−.2	.00	.96
σ_1	1	.01	.99
σ_2	1	.02	.99

of Asparouhov and Muthén (2023b). For example, the above model can also include predictors for the latent factors. The model also extends to multiple latent variables and measurement models with cross-loadings.

The Mplus input for the longitudinal alignment simulation study is given in SM Figure 9. A single latent variable is measured by $P = 3$ indicators across $T = 5$ time periods. We use the fixed alignment method to analyze the data. The specification is the same as for the MGA analysis: ALIGNMENT = FIXED. In addition to the standard MODEL statement, longitudinal alignment requires the specification of the time specific structure that is to be aligned. This is given in the MODEL T1, MODEL T2, etc, statements. These model statements specify which factor is measured at which time point and contain only latent variable measurement models, i.e., only BY statements. The alignment procedure is then applied to aligning these measurement models. We generate the data with one non-invariant loading and one non-invariant intercept. The results of the simulation for some of the parameters are given in Table 9. The bias is minimal for all parameters and the coverage is near the nominal levels.

5.4. Growth Model for Longitudinal Alignment

The longitudinal alignment model uses unrestricted mean and variance covariance structures. If the number of time points is sufficient, an additional structure can be imposed for the factor means and variances. This would make the model more parsimonious and ultimately more interpretable. For example, a latent growth curve model can be

added for the factors. Such structural models, however, are not available directly in the alignment estimation and therefore we resort to the PSEM framework. We illustrate this process with a simulation study where the aligned factors follow a linear growth curve model. In section 5.1 we showed how PSEM can be used to set up the alignment model. This is now repeated also for the longitudinal alignment model. The growth model is added as usual. Using the setup from the previous section, the longitudinal alignment growth model is given as follows. Equations (30-31) remain unchanged, i.e., measurement invariance is not assumed, while Equation (32) is replaced by the following equations

$$f_t = I + S \cdot t + \zeta_t \tag{33}$$

$$\zeta_t \sim N(0, \sigma_t) \tag{34}$$

$$I \sim N(0,1), S \sim N(m,v), Cov(I,S) = c. \tag{35}$$

$$DIFF(\nu_{i1} - \nu_{iT}) \sim ALF(0,1) \tag{36}$$

$$DIFF(\lambda_{i1} - \lambda_{iT}) \sim ALF(0,1) \tag{37}$$

Note that the random intercept has a fixed standard normal distribution, i.e., the mean is fixed to 0 and the variance is fixed to 1. This is necessary to identify the global shift and scale for the model. It plays the same role as the fixed standard normal distribution for f_1 in the longitudinal alignment model (30-32). This restriction can be replaced by other restrictions. For example, the variance of I can be estimated if the residual variance of f_1 is fixed to 1. The mean of I can be estimated if one of the indicator means is taken out of the alignment and is instead fixed to zero. Such alternative parameterizations tend to lead to overall increase in the MSE for all parameters and thus we recommend using the parameterization given by equations (30-31) and (33-37). Equations (36-37) are needed because we use the PSEM framework and not longitudinal alignment.

The Mplus input file for the longitudinal alignment linear growth model is given in SM Figure 10 and Table 10 shows the results for some of the parameters. As in the previous example, we generate the data with one non-invariant loading and one non-invariant indicator intercept. The results show that the bias is minimal and the coverage is near the nominal level for all parameters.

The linear growth model for the factors can be replaced with other structural models, such as (second order) factor analysis model or a quadratic growth model. The model can also be combined with the alignment of means growth model discussed in Section 4.2. This would allow the estimation of the factor means α_t in addition to the growth model, provided that these parameters are also included in the penalty function via the alignment priors.

6. EFA Examples

In this section we consider the possibility to estimate customized EFA models using the PSEM framework.

6.1. Estimating EFA with PSEM

First, we illustrate how to set up an EFA model estimation within the PSEM framework. In this example, we use the geomin rotation criterion

$$g(\Lambda) = \sum_{i=1}^{P} \prod_{j=1}^{M} (\lambda_{ij}^2 + \epsilon)^{1/M}, \qquad (38)$$

where P is the number of observed variables and M is the number of factors, λ_{ij} are the loadings and ϵ is a small positive number which is defaulted in Mplus to 0.01. To estimate an EFA model in the PSEM framework we want to set the penalty to $g(\Lambda)$. To accomplish that, we simply give a Geomin prior for the loading matrix

$$\lambda_{11} - \lambda_{ij} \sim Geomin(M, v). \qquad (39)$$

We illustrate this setup with a simulation study using 20 observed variables and 4 factors. Two of the factors have 5 non-zero loadings and two of the factors have 6 non-zero loading, i.e., two indicators have non-zero loadings for two factors. The Mplus input file is given in SM Figure 11. The factors are correlated in this example, i.e., we are using an oblique rotation. Table 11 contains the results of this simulation study for some of the parameters using PSEM and ESEM based EFA. The results are nearly identical, bias is minimal and coverage is near the nominal levels. The log-likelihood values and chi-square are also nearly identical, both methods yielding a 7% rejection rate, which is also near the nominal level.

In this estimation we used a prior variance $v = 0.1$ so that the penalty ratio is within the optimal range of 0.01 and 0.10. For $v = 0.1$, the penalty ratio is 0.024 and for $v = 1$ it is 0.002. Using $v = 1$ lowers the penalty weight and requires a sharper convergence criterion of 0.000001 so that the penalty is minimized completely. Using $v = 0.1$ works well with the default convergence criterion of 0.00005. This example illustrates the strong connection between the variance of the prior (the weight of the penalty function) and the convergence criterion. In general, smaller penalty ratios are likely to require sharper convergence criteria. In this example, using $v = 1$ with the sharper converge criterion and $v = .1$ or $v = 0.01$ with the default convergence criterion yield nearly identical results for the parameter estimates, i.e., the range of acceptable prior variance values is not small.

Here are some general indicators that the penalty weight might be too small: penalty ratio less than 0.01, saddle point

error message reported in the standard error computation (which is a clear indicator that the convergence is not complete and a sharper convergence criterion is needed), standard errors that are too large (when the penalty weight is too small, the model identification which is tied to the penalty minimization will be too weak and the model will appear as nearly unidentified), parameter estimates that are very close to their starting values for some of the parameters (parameters that are supposed to be identified by the penalty optimization will have nearly zero derivatives at the starting values when the penalty weight is too small), random starting values yielding mostly non-convergence (this is an indication of singular information matrix which obstructs the optimization). Some experimentation with the prior variance may be necessary when the penalty function is complex (as is the case of the rotation criterion) to ensure that a proper penalty ratio is obtained. What is not a good criterion to detect low penalty weight is the log-likelihood value. The log-likelihood value will be correct even with a very low penalty since the problem is not in the log-likelihood optimization but in the penalty optimization.

Situations when the penalty weight is too high are easier to detect. The log-likelihood value will be worse than that of the null model or models with smaller penalty weight. When the penalty is too strong, the optimization focuses more on the penalty and less on the log-likelihood which directly results in incorrect log-likelihood value. The model will provide an inferior fit to the data, and log-likelihood value that doesn't match models with smaller penalty weight and the null model. If the model is estimated with several different prior variance values, log-likelihood values that are worse should be interpreted as penalty weight being too high. In the above example, prior variance of $v = 0.001$ results in worse log-likelihood value than models with larger prior variance: $v = 1, 0.1$ and 0.01 (all of which yield the same log-likelihood value).

Figure 2 depicts the general penalty weight effect on the model. This figure applies to all PSEM models but is most well understood when the penalty is the ALF-target rotation
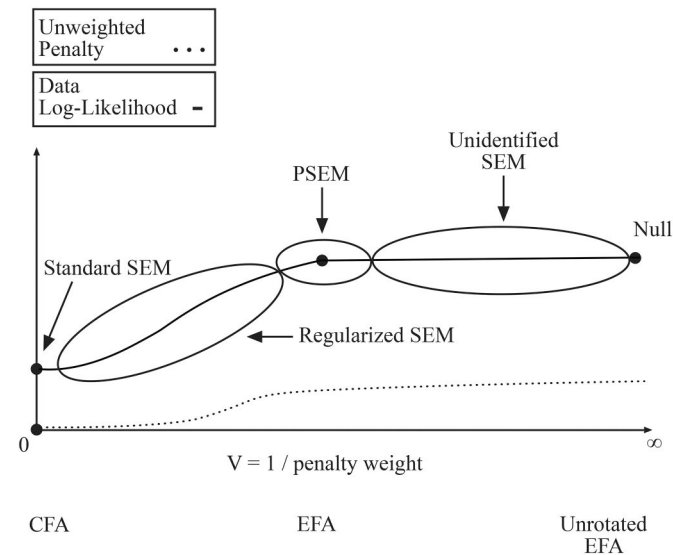
Table 11. Comparing ESEM and PSEM for EFA: Absolute bias(coverage).

Parameter	True Value	ESEM	PSEM
λ_{11}	.7	.01(.89)	.01(.91)
λ_{21}	.3	.01(.93)	.01(.94)
λ_{31}	.4	.01(.98)	.01(.97)
λ_{41}	.4	.00(.95)	.00(.96)
λ_{51}	.3	.00(.98)	.00(.97)
λ_{61}	.0	.00(.96)	.00(.96)
λ_{42}	.0	.00(.98)	.00(.98)
λ_{52}	.6	.00(.97)	.00(.98)
λ_{62}	.7	.01(.96)	.01(.97)
ψ_{12}	.4	.02(.96)	.02(.94)

Figure 2. Penalty weight effect on models.

Table 12. Hierarchical EFA.

Parameter	True Value	Abs. Bias	Coverage
$\lambda_{11,1}$	.7	.00	.94
$\lambda_{21,1}$	1.3	.00	.93
$\lambda_{31,1}$	.8	.00	.95
$\lambda_{41,1}$	.8	.00	.96
$\lambda_{51,1}$	.3	.00	.95
$\lambda_{61,1}$	.0	.00	.97
$\lambda_{42,1}$	.0	.00	.98
$\lambda_{52,1}$	.6	.00	.96
$\lambda_{62,1}$	.7	.00	.96
$\lambda_{11,2}$	.4	.01	.98
$\lambda_{21,2}$	.5	.01	.99
$\lambda_{31,2}$	.5	.02	.95
$\lambda_{41,2}$	.6	.02	.98

discussed earlier. When the penalty weight is infinity ($\nu = 0$), the penalty is minimized before the data log-likelihood. The penalty will then be minimized to zero, which essentially converts the penalty to a fixed parameter constraint. In the EFA-target case, the targets become fixed to 0 and the model becomes the CFA model. For a general PSEM model, $\nu = 0$ yields a standard SEM model, where the penalty is converted to a parameter constraint. As the variance ν increases, the data log-likelihood increases. That region of the plot corresponds to regularized SEM models where prior variance still has an impact on the model. As the variance increases even more we reach the area of the plot where the log-likelihood stops increasing. This is precisely where the PSEM model lies. The data log-likelihood is completely optimized at that point and PML optimization can turn to minimizing the penalty and mimic the conditional two-stage optimization as in the EFA estimation. As the variance increases even more, numerically the penalty would be too small to produce the necessary model identification and thus we will obtain an unidentified model. The right part of this figure is the null model of the PSEM model where identification is provided by manually fixing the unidentified dimensions, which corresponds to the unrotated EFA model. Figure 2 shows how varying the penalty weight provides a continuum of models that connects standard SEM, PSEM, and the null model. The unweighted penalty is also depicted in Figure 2. This quantity can also be of interest in certain situations as it provides a comparative measure for how much the model constraint of the standard SEM model must be relaxed so that the data log-likelihood can reach the log-likelihood of the null model. For example, in multiple group alignment, the unweighted penalty reflects the amount of measurement non-invariance. By comparing the unweighted penalty for the PSEM and the null model we can determine the reduction in non-invariance obtained by alignment.

Input file for PSEM-EFA estimation for non-simulation studies, i.e., estimation with real data, is provided in SM Figures 12.

6.2. Hierarchical Exploratory Factor Analysis

In EFA models, it is often the case that the factors are highly correlated. This prompts the question if it is possible to estimate a second order factor analysis model using the EFA factors as indicators. One possible solution has been proposed in Morin and Asparouhov (2018) based on the two stage estimation process ESEM-within-CFA. The PSEM methodology offers a simple one stage estimation. The model can be described as follows. Suppose that Y is a vector of dependent variables measuring a vector of latent variables F via an EFA model

$$Y = \nu + \Lambda_1 F + \varepsilon. \qquad (40)$$

The exploratory factors F are then used as indicators for a secondary EFA model measuring a vector of second order factors η

$$F = \Lambda_2 \eta + \xi. \qquad (41)$$

In the above equations all parameters in Λ_1 are free parameters as the EFA model specifies. The parameters in Λ_2 are also free parameters (secondary EFA model) or Λ_2 can have a CFA like structure (secondary CFA model). For identification purposes ξ consists of standard normal random variables, i.e., the residual variances of F are fixed to 1. The variances of η are also fixed to 1. To estimate this model with PSEM, the geomin prior is given for Λ_1 and Λ_2 as in the previous section. If the secondary model is a CFA model, only Λ_1 prior is used.

We should note here that the hierarchical EFA (HEFA) model described above is an example of a rotation that is neither oblique or orthogonal. The variance covariance of the first order EFA factors is a structured variance covariance, i.e., the factors are not independent and are not completely unconstrained. Such structured rotation is not possible with the traditional EFA methodology and thus PSEM offers a unique solution.

We illustrate the performance of this method with a simulation study with $P = 20$ observed variables, 4 first order exploratory factors, and 1 second order factor. Because we use only one second order factor, we can interpret the second order model as a CFA model and use only the rotation criterion for the first order EFA. The setup for this simulation study is similar to the one used in the previous section. Most of the indicators have only one non-zero loading. Only two variables load on 2 of the factors. The Mplus input file for this simulation study is given in SM Figure 13. The results for some of the parameters are given in Table 12. The bias in the parameter estimates is minimal and the coverage is near the nominal levels.

In this simulation study, the prior variance of $\nu = 0.01$ is selected to produce an acceptable penalty ratio of 0.02. In addition, we use 0.001 as the small positive number in the geomin prior. Lowering this number reduces the bias in the parameter estimates for more complex models.

The HEFA model used in this simulation, i.e., HEFA model with one second order factor, is similar to the bi-factor EFA model. To see this similarity, one can combine equations (40-41) into the following equation

$$Y = \nu + \Lambda_1 \xi + \Lambda_1 \Lambda_2 \eta + \varepsilon \qquad (42)$$

which resembles a bi-factor model. The second order factor η becomes the general factor and the residuals ξ become the specific factors. The model is not precisely a bi-factor EFA

model because the general factor loadings are structured/-constrained. Therefore the HEFA model with one second order factor is nested within the bi-factor EFA model. The HEFA model has fewer parameters because the second order factor has fewer loadings as it loads only on the first order factors rather than all observed variables as in the bi-factor EFA model. Another important difference between the two models is in the interpretation. With the HEFA model, the common factor is common to the latent measurements, rather than all observed variables.

6.3. Bi-Factor EFA with More than One General Factor

This model was suggested by Herb Marsh in personal communication. The idea is that specific factors can be used to model high correlations between items that are similar in nature, while general factors are true latent features that are measured by all observed data. The role of the specific factors is secondary to some extent and is only there to account for additional correlation caused by similarity in the items. From this point of view, there is no reason to limit the number of general factors to 1. Here we define the generalized bi-factor EFA model with more than one general factor as follows

$$Y = \nu + \Lambda_1 G + \Lambda_2 F + \varepsilon \quad (43)$$

where G is a vector of general factors and F is a vector of specific factors. All parameters in Λ_1 and Λ_2 are free, the correlation matrix $Var(G)$ is an unrestricted correlation matrix, $Var(F)$ is the identity matrix and Cov(F,G)=0. This is another example of a non-oblique and non-orthogonal rotation. The rotation is oblique for G but orthogonal for F.

Recall that the geomin rotation criterion for the standard bi-factor EFA with one general factor excludes the general factor loadings. With more than one general factor, however, it would not be possible to use such a rotation criterion because the loading structure Λ_1 must be rotated and identified. To estimate the above model with PSEM, we use the full geomin rotation function using the full loading matrix which includes both Λ_1 and Λ_2. Thus, the difference between the generalized bi-factor model and the standard EFA model is in the factor covariance structure: the general factors can be correlated while the specific factors must be independent. Note, however, that if a general factor happens to also be nearly independent of the remaining general factors, there is the possibility that the general factor and a specific factor will switch places and that may cause interpretation issues.

We illustrate the above model with a simulation study using $P = 20$ variables, 2 general factors that each load on 10 different variables. We also include 2 cross-loadings for the general factors, i.e., two of the variables load on both general factors. In this model there are 3 specific factors, 2 factors load on 6 variables, while the third factor loads on 8 variables. There is one cross-loading for the specific factors as well. The Mplus input file for this simulation study is given in SM Figure 14 and the results for some of the model

Table 13. Generalized bi-factor EFA.

Parameter	True Value	Abs. Bias	Coverage
$\lambda_{1,1,1}$	.7	.00	.96
$\lambda_{2,1,1}$	.3	.00	.96
$\lambda_{3,1,1}$	.4	.01	.97
$\lambda_{10,1,1}$	1	.00	.97
$\lambda_{11,1,1}$	.2	.01	.97
$\lambda_{12,1,1}$	.3	.00	.97
$\lambda_{10,2,1}$	0	.00	.97
$\lambda_{11,2,1}$	1	.00	.91
$\lambda_{12,2,1}$	1	.00	.94
$\lambda_{3,1,2}$	0	.01	.96
$\lambda_{4,1,2}$	0	.00	.99
$\lambda_{5,1,2}$	.6	.01	.97
$\lambda_{3,2,2}$	1	.01	.95
$\lambda_{4,2,2}$	1	.00	.93
$\lambda_{10,2,2}$	0	.00	.97

parameters are given in Table 13. The bias in the estimates is minimal and the coverage is near the nominal levels.

It should be noted here that the generalized bi-factor model is easier to estimate than the HEFA model described in the previous section. This means that the model can be estimated with smaller samples. This estimation is essentially a standard EFA estimation with the addition that some of the correlations between the factors are constrained to be zero, i.e., estimating this model is unlikely to be substantially harder than estimating a standard EFA model.

In principle, it is possible in PSEM to add one rotation function for the general factors and a separate rotation function for the specific factors. A simulation study with this approach revealed that the model can not be recovered, i.e., the rotation of the specific and the general factors must be joint.

In the above model, both Λ_1 and/or Λ_2 can be converted to structured CFA loading matrices where some of the parameters are fixed to zero, i.e., the general factor can remain EFA, while the specific factors become CFA, and the reverse. Such a model can be estimated with both PSEM and ESEM.

6.4. Scalar and Metric Invariant EFA with Orthogonal Rotation in Multiple Group and Longitudinal Settings

In this section we describe a problem currently existing with orthogonal EFA estimation where loadings are held equal across two or more EFA model components. This situation arises with metric and scalar invariant models in multiple group settings where EFA loadings are held equal across groups. It also arises in longitudinal studies when we have repeated measurements for a set of EFA indicators across time and the EFA loadings are time invariant. Here we illustrate how the PSEM methodology can be used to resolve this issue in the case of scalar invariant multiple group EFA, however, the proposed approach extends also to longitudinal settings as well as metric invariant models.

Consider a multiple group scalar EFA model. The standard setup for such a model is that the loadings and indicator intercepts are held equal across groups. With oblique rotation, in the reference group the factor means are fixed to zero, the factor variances are fixed to 1, while the factor

covariances are estimated as free parameters, i.e., in the reference group, the model is the same as the single group analysis. In all other groups, factor means, variance and covariances are estimated. This estimation is implemented in the Mplus ESEM framework and is performed as follows. An unrotated scalar model is estimated first and then the invariant loading matrix is rotated for all groups. For the unrotated scalar model, in the reference group the factor distribution is fixed to the standard normal distribution, while in all other groups it is unconstrained normal distribution. When the unrotated model is rotated, in the reference group we obtain a zero mean full correlation matrix for the factor distribution, and in all other groups the factor distribution is again unconstrained.

Next we consider the multiple group scalar EFA model with orthogonal rotation. Traditionally, the orthogonal and the oblique EFA use the same unrotated model. A general reason for that is to ensure that the model fit is not affected by the type of the rotation. The only thing that changes between the orthogonal and the oblique rotation is the type of rotation we are allowed to use in the estimation. If we apply orthogonal rotation to the unrotated model described above, the resulting factor distribution in the non-reference groups would not be orthogonal because the unrotated factor distribution is unconstrained. The factor distribution will be orthogonal only in the reference group because the unrotated factor distribution in that group is the standard normal distribution. Even if we change the unrotated distribution to orthogonal in the non-reference groups (which is not desirable because that will affect the fit of the model), the rotated distribution will be orthogonal only if the factor variances in the unrotated and rotated model are identical across factors (this kind of constraint appears unreasonable). If we constrain the factor variance/covariance to be the identity matrix in all groups in the unrotated model we will obtain rotated orthogonal model for all groups but that is undesirable because the factor variance covariance will be identical in all groups in the rotated model (it will be the identity matrix). We therefore conclude that the standard EFA methodology can not estimate multiple group scalar invariant EFA with orthogonal rotation where the factors are orthogonal in all groups. Indeed, the Mplus estimation of such a model will show that only the reference group has orthogonal factors.

The above dilemma presents itself for example when we want to estimate a scalar multiple group bi-factor EFA model. Orthogonal rotation is needed to ensure that the general and the specific factors are uncorrelated. It can also become an issue for those situations where the reference group is of substantial size while the non-reference groups are not. Not being able to constrain the factor covariances in the non-reference groups to zero will interfere with the benefit of scalar invariance where information from the big reference group transfers to the smaller groups.

The PSEM methodology can be used to resolve the above problem. We simply must specify uncorrelated factors in all groups, allow the factor means and variances to be estimated as group-specific, and specify geomin prior for the

invariant loadings. We illustrate this approach with a simulation study based on a 2-group, 2-factor model with a total of 6 observed variables, where each factor is measured by 3 different main indicators. We also introduce one cross-loading. The Mplus input for this simulation study is given in SM Figure 15 while the results are given in Table 14. Here λ_{ij} is the loading for variable i on the j-th factor which is group invariant. Factor mean and variance for the j-th factor in group g are denoted by α_{jg} and σ_{jg}.

We used a small sample size in this example to provide a realistic setting when sharing information across groups is particularly beneficial. The results indicate that the parameter bias is minimal and coverage is near the nominal levels. The average chi-square value for this example is 21.9 and with 20 degrees of freedom the rejection rate is 7% which is near the nominal level as well.

6.5. Partial Invariance EFA

In multiple group EFA analysis, scalar invariance is most desirable, however, it is often unrealistic particularly when the sample size is large. The AESEM framework, Asparouhov and Muthén (2023b) provides one solution to this problem where EFA loading structures are aligned between the groups. This means that through optimizing the alignment loss function and the rotation criterion, most parameters will be estimated to be approximately equal across the groups and only those that are truly non-invariant will be estimated as different parameters. This framework is exploratory in nature not just for the factor formation but also for the identification of non-invariant parameters. Here we discuss a slightly different approach: exploratory for the factor formation but confirmatory for the non-invariance of parameters. The AESEM method also has a different disadvantage. It is often the case that invariant parameters are estimated as approximately invariant but they are not exactly the same. To some extent, that causes somewhat of an uncertainty whether the parameters are truly the same or not. The PSEM approach that we offer here, produces invariant parameters that are exactly the same. Furthermore, the AESEM approach can be used as a preliminary analysis for the PSEM method. Using the AESEM method, the non-invariant parameters can be identified. Then, the PSEM partial invariance EFA can be used as a confirmatory, more parsimonious version of the AESEM model, where parameters that are identified as invariant are actually held equal to each other.

Table 14. Scalar invariant orthogonal EFA.

Parameter	True Value	Abs. Bias	Coverage
λ_{11}	1	.01	.95
λ_{41}	0	.00	.98
λ_{12}	.3	.03	.98
λ_{42}	1	.01	.96
α_{12}	.4	.02	.95
α_{22}	.9	.02	1.0
σ_{12}	1.3	.01	.90
σ_{22}	1.4	.04	.97

The partial invariance PSEM-EFA model simply allows us to specify group specific loadings and intercepts, where such are needed, while still retaining the general rotation framework. Non-invariant loadings can be specified even for all indicators for a particular factor, as long as we do not attempt to estimate the factor variance at the same time. The method can be described as follows. An EFA model is used for the reference group. For all other groups, the loadings and intercepts are held equal to those in the reference group, except for a subset of those parameters that are specified as group specific non-invariant parameters. The geomin prior is specified for the loading matrix in the reference group.

We illustrate this method with a simulation study using a 2-group, 2-factor EFA with 6 indicators, 3 main indicators per factor. We use one cross-loading in the EFA. Furthermore, we generate the data with one non-invariant loadings and one non-invariant intercept. The estimated model is specified with the same non-invariance structure, i.e., the estimated model is a scalar invariant EFA model with the exception of one non-invariant intercept and one non-invariant loading. Because scalar invariance holds for most of the parameters in the model, we can estimate group-specific factor means, variances, and covariances. In the reference group, for identification purposes, the factor means are fixed to 0 and the factor variances are fixed to 1.

The input file for this simulation study is given in SM Figure 16 and the results for some of the parameters are given in Table 15. Here λ_{ij} denotes group invariant loading for variable i on the j-th factor, while non-invariant loadings are denoted by $\lambda_{ij,g}$. Similarly, ν_i denotes a group invariant intercept for variable i, while $\nu_{i,g}$ denotes a non-invariant intercept. Factor mean and variance/covariance are denoted by $\alpha_{j,g}$ and $\sigma_{ij,g}$. The bias of the parameter estimates is minimal and the coverage is near the nominal levels.

We should also note here that the intercept invariance can be estimated within the ESEM framework, in a similar way. Simply estimating a particular set of intercepts as non-invariant. The ESEM framework can not, however, provide partial invariance for the loadings.

6.6. Comparison to Regularized EFA and Sparse EFA

Several penalized methods have been proposed recently for EFA estimation, see Hirose and Yamamoto (2014, 2015), Trendafilov et al. (2017) and Scharf and Nestler (2019) among others. These methods are fundamentally identical to PSEM, i.e., the penalized EFA methods described above. The unidentified EFA model is rotated by the penalty function. We also pointed out here that the penalty function and the rotation criterion are one and the same. Thus, if the penalty function is specified as lasso, the estimation becomes equivalent to the component loss function rotation discussed in Jennrich (2006), i.e., the lasso rotation is the same as the component loss function method when using absolute loading value. This understanding can be useful in choosing penalty/rotation. Generally, much more information is available regarding the performance of the various rotation

Table 15. Partial invariance EFA.

Parameter	True Value	Abs. Bias	Coverage
λ_{11}	1	.00	.98
λ_{41}	0	.01	1.0
λ_{61}	0	.02	.98
λ_{12}	.3	.03	.92
$\lambda_{42,1}$	1	.01	.97
$\lambda_{42,2}$	.5	.01	.89
λ_{62}	1	.02	.94
$\nu_{6,1}$	0	.00	.99
$\nu_{6,2}$	.5	.00	.97
$\sigma_{12,1}$	.3	.01	.94
$\sigma_{11,2}$	1.3	.04	.98
$\sigma_{12,2}$	.1	.00	.95
$\sigma_{22,2}$	1.4	.08	.94
$\alpha_{1,2}$	.4	.01	.94
$\alpha_{2,2}$	.9	.01	.97

criteria traditionally used with EFA, but the regularized/penalized EFA is more flexible in terms of penalty function. In the above examples we used the Geomin rotation as the penalty function. The Geomin method has a long history and a consistent performance, see Browne (2001). Lasso rotation may also be suitable in many situations. However, a word of caution is necessary here. Penalized/Regularlized EFA is so flexible that it may inadvertently produce bad outcomes. The penalties used for rotations should be only those that have documented performance. As an example, Quartimax performs well only with orthogonal rotations and it has poor performance with oblique rotations. Penalty functions that resemble Quartimax (fourth power of loadings) should be used only when appropriate.

6.7. The Native Model Rotation

PSEM-ESEM modeling framework is so flexible that any kind of structural model can now be formulated for the exploratory factors, i.e., models that have never been estimated before. This however leads to the rise of a new concept. In some models, the structure of the exploratory factors is so restrictive that we no longer need a rotation criterion. The restrictions on the exploratory factors can serve as the rotation criterion, i.e., the model has sufficient structure that unrestricted loading matrix can be estimated without any rotation, penalty, PSEM or ESEM frameworks. None of the models we described above are such models. The simplest example is as follows. A two factor model is estimated where each factor has its own separate predictor. The EFA rotation in this case identifies just two parameters. The structure of the model implies that there are two cross regression parameters that are fixed to zero. These two restrictions are enough to replace the rotation criterion and the model can be estimated as a standard SEM model even though all loadings are free. If we estimate such a model as a PSEM-ESEM model, the data log-likelihood will continue to increase as we decrease the penalty weight down to zero (increase the variance to infinity) and it will not plateau as described in Figure 2. Another interesting example for a model with native rotation is as follows. Consider a 2-factor longitudinal EFA model where each of the two factors has an auto-regressive model across time (but not cross-lag

regression). The fact that the cross-lags are fixed to zero can act as a rotation criterion, i.e., the model can be estimated as an SEM model without any rotation and not as a PSEM-ESEM model. Models with native rotation also are problematic in their construction of the null model. The unrotated EFA which we generally accept as the null model in that case will have a worse log-likelihood than the native rotation SEM model.

7. Exploratory Latent Growth Models

Tucker (1958, 1966) proposed growth curve modeling that is exploratory in nature. Instead of using the times of observation as the loading parameters, all loadings are estimated. The loadings are also rotated to the most interpretable solution as in EFA. This model has subsequently been referred to as Tuckerized Growth Curve Analysis (TGCA).

Earlier works on this topic were based on principal component analysis instead of factor analysis, see Arbuckle and Friendly (1977). Most recently, Grimm et al. (2013) used the factor analysis based ESEM framework of Asparouhov and Muthén (2009), to estimate Exploratory Latent Growth Models (ELGM). It has been pointed out that standard rotation methods which attempt to rotate the estimated loading matrix to simple structures with minimal cross loadings is not appropriate for such analysis. Using standard rotation criteria such as geomin for ELGM leads to rotation of the growth factors that doesn't resemble a growth model like (20). Geomin would yield time specific processes in its attempt to minimize cross-loadings. Bi-factor EFA might be somewhat more appropriate as it would be able to extract one main trajectory that is complemented by some time specific deviations but this approach would be appropriate only in some situations.

One of the principal aspects of TGCA is the factor rotation that produces the most interpretable growth curves. It has been proposed that the best rotations are those that produce smooth curves, curves with all positive loadings, and curves with asymptotes. A variety of rotation methods have been proposed, however, it has become clear that it would be difficult to find a universal rotation criterion that can be suitable for all situations. A rotation criterion that favors linear growth is unlikely to be usable for situations where exponential or logarithmic growth is needed. Furthermore, none of these criteria have been implemented in structural equation modeling software which limits the practical impact of the methods. The PSEM framework allows for unlimited customization of the rotation criteria and thus appears to be the perfect tool for studying ELGM/TGCA. After a brief review of ELGM and issues related to it, PSEM-ELGM is presented and illustrated.

7.1. The ELGM Model and Its Corresponding Null Model

Suppose that Y_t is a variable observed across a population of individuals at time $t = 1, ..., T$. The ELGM model with m growth factors is defined as

$$Y_t = \sum_{j=1}^{m} \lambda_{tj} \eta_j + \varepsilon_t, \qquad (44)$$

where η_j are individually specific latent variables. The vector $\eta = (\eta_1, ..., \eta_m)$ is assumed to have a normal distribution with mean $\alpha = (\alpha_1, ..., \alpha_m)$ and a correlation matrix Ψ, i.e., we assume that $Var(\eta_j) = 1$. The residual variables ε_t are assumed to have $N(0, \theta_t)$ distribution. The loadings parameters λ_{tj} are to be estimated. We denote the matrix of all loading parameters by Λ and the diagonal matrix with all residual variance parameters by Θ. In matrix form the above equation can be written as

$$Y = \Lambda \eta + E \qquad (45)$$
$$\eta \sim N(\alpha, \Psi), E \sim N(0, \Theta). \qquad (46)$$

The parameters of this model are Λ, α, Ψ and Θ. The model is a hybrid combination of the EFA and the standard growth curve models. If we remove the m parameters in α and replace them with T intercept parameters for the Y vector, we obtain the standard EFA model. Thus, the ELGM is essentially an EFA model with mean structure conflated with the variance covariance structure. Also, if we fix the loading matrix to the time scores, we obtain the standard latent growth models. If $m = 2$, $\lambda_{t1} = 1$, and $\lambda_{t2} = t$, the model becomes the standard linear growth model.

The above model is clearly unidentified and it is thus essential to determine the null model. That is, we need to find an identifiable model that has the same data fit and log-likelihood value as (45-46). This is essential for multiple reasons. In practical applications it is important to separate the process of rotation to the most interpretable model and the fit of the model. For example, if the null model has poor fit, pursuing various rotation techniques will not resolve the problem. Furthermore, when using PSEM, estimating the null model provides a baseline for the penalized model. The log-likelihood values for the two models must be nearly identical. That ensures us that the penalty function is set properly, i.e., it is not weighted too heavily as to unintentionally damage the data fit in our attempt to improve model interpretation by rotation. The null model also provides the correct degrees of freedom needed for testing purposes.

For the ELGM model, the EFA theory can be used to determine the null model. We can ignore the mean structure and focus on the loading matrix Λ and the factor correlation matrix Ψ. It is well known that an oblique rotation matrix H exists that can rotate Ψ into the identity matrix and Λ into a loading matrix with zeros above the main diagonal. This is the same rotation that can rotate the EFA model into the unrotated EFA solution used in the EFA estimation and reported in TECH1 in Mplus. Thus, H satisfies the following

$$HH^T = \Psi$$
$$\Lambda_0 = \Lambda H$$
$$\lambda_{0, tj} = 0, \text{ for } j > t.$$

If $\eta = H\eta_0$, then model (45-46) becomes

$$Y = \Lambda_0 \eta_0 + E \qquad (47)$$
$$\eta \sim N(\alpha_0, I), E \sim N(0, \Theta), \qquad (48)$$

where $\alpha_0 = H^{-1}\alpha$. Thus we have eliminated the $m(m-1)$ unidentified parameters from model (45-46). Model (47-48)

is identified because it is nested within the unrotated EFA model. This model also has the same log-likelihood as the general ELGM model (45-46). We conclude that the ELGM null model is model (47-48). We can also now compute the degrees of freedom for the ELGM model. The unrotated ELGM model (47-48) has $Tm - m(m-1)/2 + m + T$ parameters: Λ_0, α_0 and the diagonal of Θ. Thus the degrees of freedom for this model is $T(T+1)/2 - Tm - m + m(m-1)/2$, which is exactly $T - m$ more degrees of freedom than the EFA model.

The above observations also serve as the foundation of the ELGM estimation proposed in Grimm et al. (2013). The top part of SM Figure 17 shows the Mplus input file for estimating the geomin rotated ELGM via ESEM. The middle part of SM Figure 17 shows the unrotated ELGM null model estimated via CFA with zero loadings in the upper right-hand part of Λ_0.

7.2. ELGM Model with Auto-Correlation

A growth curve model is essentially a time-series model. Therefore the model should include at least a minimal attempt to model auto-correlation. This is particularly the case for ELGM models. If auto-correlations are not included in ELGM, it will likely lead to needing many more factors than necessary to fit the data. Using the height data discussed in Grimm et al. (2013), we found that a single auto-correlation parameter can improve the data fit by as much as adding 3 additional factors. Adding auto-correlation to the model is not related to the rotation of the curve model. It only replaces the Θ diagonal matrix with a non-diagonal matrix. Thus, adding auto-correlation to the ELGM or the null ELGM will provide the same data fit improvement. A simple way to introduce auto-correlation is to use the Residual Structural Equation Modeling (RSEM) framework discussed in Asparouhov and Muthén (2023a). In that framework a secondary structural model is constructed for the residuals of the primary structural model. To introduce auto-correlation in ELGM we replace the assumption of diagonal residual variance covariance matrix Θ with the following RSEM, for $t > 1$

$$\varepsilon_t = r\varepsilon_{t-1} + \varepsilon'_t \qquad (49)$$

$$E' = (\varepsilon_1, \varepsilon'_2, ..., \varepsilon'_T) \sim N(0, \Theta). \qquad (50)$$

In this model, Θ is also diagonal but the variance covariance for E is not and it follows an autoregressive pattern. The auto-regressive coefficient r in (49) can also be time-specific. In our examples, however, we use a time-invariant auto-regressive coefficient. Thus, a single model parameter is added to incorporate an autoregressive error structure.

In the Mplus language, the residual for a variable Y is referred to as Y. The third input file in SM Figure 17 illustrates the auto-regressive ELGM null model using the hats notation. The auto-regressive ELGM model can also be estimated by adding the autoregressive model to the first input file in SM Figure 17.

7.3. Restrictive Mean Structure

The ELGM model, just like the standard growth model, imposes a very restrictive constraint on the means of the observed variables. Small deviations from that constraint may lead to a model rejection because mean parameter estimates typically have smaller standard errors than parameters related to variance covariance modeling. The ELGM model is nested within the EFA model with the same number of factors and the likelihood ratio test for the two models specifically tests the mean constraint imposed by ELGM. This test is independent of the rotation in ELGM and can be conducted prior to curve rotation. The test can be performed with rotated or unrotated models. If the mean structure is rejected, the PSEM modeling framework offers a solution outlined in Section 4.2 for the standard growth model. For the ELGM case, Equation (45) is replaced by

$$Y = \nu + \Lambda\eta + E \qquad (51)$$

$$\nu \sim ALF(0,1). \qquad (52)$$

The parameter vector ν offers time specific deviations from the mean structure imposed by ELGM. Since the factor means α are also included in this model, we use ALF(0,1) penalty for every parameter in ν. This prior will keep the mean structure deviations to a minimum.

7.4. Best Rotations for ELGM

It has been proposed previously that the most appropriate rotations for ELGM should be rotations that produce smooth curves. That is, $\lambda_j(t) = \lambda_{tj}$ is a smooth function of t. For example, Arbuckle and Friendly (1977) define as a measure of smoothness

$$S = \sum_t (\lambda_j(t) - \lambda_j(t-1))^2.$$

The best rotation for the ELGM is then selected as the one that minimizes S. This kind of estimation can easily be accommodated in the PSEM framework. The ELGM model (45-46) is estimated with $N(0,1)$ priors given for all $\lambda_j(t) - \lambda_j(t-1)$ parameters.

Other suggestions that have been mentioned in the literature are rotations to all positive loadings and to curves that have an asymptote. Having all positive loadings can also be modeled with PSEM but it is somewhat difficult to justify. First, it often performs poorly in simulation studies. Second, even if $Y(t)$ is monotonic and the distribution of η_j is positive, monotonicity of $\sum_{j=1}^{m} \lambda_j(t)\eta_j$ does not imply monotonicity or positivity for individual curves $\lambda_j(t)$.

Next, we illustrate how to construct rotation criteria in PSEM for four different curves: zero, constant, linear, and quadratic curves. If we want $\lambda_j(t)$ to be approximately 0, we use ALF(0,1) prior for $\lambda_j(t)$. If we want $\lambda_j(t)$ to be approximately constant we setup ALF(0,1) prior for DIFF($\lambda_j(t)$), i.e, for every pair of loadings in the set, we minimize the distance. This yields a penalty function

$$P = \sum_{t_1 < t_2} f(\lambda_j(t_1) - \lambda_j(t_2))$$

where f is the alignment loss function given in (11).

If we want $\lambda_j(t)$ to be approximately a linear function, we use the following observation: $\lambda_j(t) - \lambda_j(t-1)$ is constant, which can be viewed as an approximation to the first derivative of $\lambda_j(t)$. Thus, we can use an ALF(0,1) prior for DIFF($\lambda_j(t) - \lambda_j(t-1)$). The penalty function in that case is

$$P = \sum_{t_1 < t_2} f(\lambda_j(t_1) + \lambda_j(t_2 - 1) - \lambda_j(t_2) - \lambda_j(t_1 - 1)).$$

If we want $\lambda_j(t)$ to be approximately a quadratic function, where the second order derivative $\lambda_j(t+1) + \lambda_j(t-1) - 2\lambda_j(t)$ is constant, we use an ALF(0,1) prior for DIFF($\lambda_j(t+1) + \lambda_j(t-1) - 2\lambda_j(t)$). The penalty function in that case is

$$P = \sum_{t_1 < t_2} f(\lambda_j(t_1 + 1) + \lambda_j(t_1 - 1) - 2\lambda_j(t_1) - \lambda_j(t_2 + 1)$$
$$- \lambda_j(t_2 - 1) + 2\lambda_j(t_2)).$$

Other penalties can be constructed similarly. Most importantly, the above constructions can be used in piecewise fashion. For example, $\lambda_j(t)$ can be linear from time 1 to time T_0, and it can be another linear or quadratic function from time T_0 to time T. That requires constructing a penalty function for $\lambda_j(1), ..., \lambda_j(T_0)$ and then constructing a separate penalty function for $\lambda_j(T_0), ..., \lambda_j(T)$.

Note here that the ELGM model contains m smooth curves $\lambda_j(t)$. A penalty function should be constructed for each of these curves separately by adding priors for each column of the loading matrix. The total penalty function is the sum of these m penalty functions.

There are several conditions that must be satisfied by a rotation criterion to be useful in practical applications. First, the PSEM model estimation should converge. Second, the results obtained from the rotated estimation must be close to what the criterion is designed for. Any large deviation should be considered as a poor choice of rotation. Third, the standard errors should remain within the range of the standard errors of the null model. If the standard errors become too big, the rotation/penalty function is not sufficient to identify the model well. Fourth, the rotation should perform well in simulation studies.

7.5. ELGM Model Identifiability Limitations

Consider the ELGM model with $m = 2$ factors, where both $\lambda_1(t)$ and $\lambda_2(t)$ are linear. The model can be described as follows

$$Y_t = \lambda_1(t)\eta_1 + \lambda_2(t)\eta_1 + \varepsilon_t. \tag{53}$$

Consider now a rotation of this model $\eta = H\eta_0$

$$\eta_1 = H_{11}\eta_{0,1} + H_{12}\eta_{0,2}$$
$$\eta_2 = H_{21}\eta_{0,1} + H_{22}\eta_{0,2.}$$

The rotated model becomes

$$Y_t = (H_{11}\lambda_1(t) + H_{21}\lambda_2(t))\eta_{0,1} + (H_{12}\lambda_1(t) + H_{22}\lambda_2(t))\eta_{0,2} + \varepsilon_t.$$

In this rotated model, both loading curves $H_{11}\lambda_1(t) + H_{21}\lambda_2(t)$ and $H_{12}\lambda_1(t) + H_{22}\lambda_2(t)$ are also linear curves.

Therefore the rotated model is not distinguishable from the original model (53). If we simply impose a rotation criterion corresponding to two linear curves for the above model, we will not be able to identify the model properly. Furthermore, neither one of the two models can be claimed to be more interpretable than the other model. Identification in such a situation will require removing parameters from the model or altering the rotation criterion.

The same thing applies if one curve is constant and the second curve is linear. That is because adding a constant curve to a linear curve still produces a linear curve. In a standard linear growth model, where we do have a constant loading function for the random intercept and a linear function for the random slope, the identification is resolved by fixing the first loading for the slope to 0. Similarly, if one curve is linear and the second is quadratic, the model is still not identifiable. An example of curves that are identifiable is this: a linear curve and a curve that is piecewise constant and quadratic.

In practical examples, a two-linear function ELGM model may actually appear identified. That is because the functions are likely not perfectly linear. Such imperfections may force the penalty function minimization to produce a proper minimum over all possible rotations. In such cases, however, replicability is most likely questionable and properly conducted simulation studies will likely reveal the quality of the estimation.

Overall, selecting a rotation criterion for ELGM, remains a difficult task. Simulation studies should be conducted with every application to ensure that the rotation criterion works as expected.

7.6. ELGM Model Limitations Due to Degrees of Freedom

Consider the case of ELGM $m = 2$ again. The rotation of the model is determined uniquely by $m^2 = 4$ parameter constraints among the Ψ and Λ parameters. Two of the constraints have already been set because the diagonal values of Ψ are fixed to 1. If we use a linear growth style modeling, the first loading for the slope factor is naturally fixed to 0. Therefore only one parameter remains to be determined. Because we are identifying only a single parameter with the PSEM penalty, it is imperative that the ELGM rotation is simple and intuitive. Constructing complicated rotation criteria can not be justified by the identification of a single parameter.

It is helpful to consider how the SEM framework is used to model generalized curves, see Sterba (2014). Linear growth models in the SEM framework can be made more flexible and similar to ELGM by estimating most of the loadings instead of fixing them to the time score, i.e., imposing as few restrictions as possible. To identify a SEM generalized linear growth model, the most intuitive way is to fix the variance of the factors to 1, fix the first loading of the slope to 0, and the fourth restriction would come from holding equal the first and the last loading of the random intercept. Recall that in the standard linear growth model,

all loadings of the random intercept are held equal when the random intercept variance is fixed to 1. Using this as a guide, the rotation for the ELGM linear growth model should consist of an ALF(0,1) prior for the first slope loading and a DIFF prior for the random intercept loadings. The assumptions of this ELGM rotation are minimal and we can let the data drive the shape of the curves. Furthermore, the PSEM model will produce better estimates than the corresponding SEM model due to possible misspecifications in the SEM model. For example, if the first and the last loadings are held equal but they are not, the SEM parameters will be biased. Note that holding the first and the last random intercept loadings equal in SEM is optimal in the sense that it provides a measurement model for the random intercept with the most orthogonal measurements. At the same time, the more distant the observations are, the more likely it is that a change in those loadings had occurred.

7.7. Sequential Estimation of the ELGM Model

We argue here that it is important to estimate a sequence of ELGM models. First begin with $m = 1$, then gradually increase m to the needed dimensions. The first curve in the model fitting corresponds to the main principal component. It is the most important predictor for the Y_t trajectory. The second curve in the ELGM model should be viewed as modeling the residual trend, a trend that was not explained by the first curve, etc. Using such a sequential approach can help us identify the shape of the curves and set proper rotation criteria.

7.8. The Simplest ELGM Pattern

Here we discuss an issue that mostly arises in simulation studies and not in practical applications. The issue is that in some PSEM simulation studies, a perfectly good rotation criterion does not recover the generating values, i.e., the estimates appear to be biased. Here we outline a procedure to check if the estimates are truly biased or a simpler pattern with the same data fit has been found by the PSEM optimization. In a typical PSEM simulation study, we want to see to what extent the method can tolerate deviations from expectations. As an example, consider the case of $m = 2$ where the first/main curve is set to be a random intercept, i.e., a curve with constant loadings. If we generate data using a model where the loadings are not exactly constant, we expect PSEM to produce unbiased estimates nevertheless. There is a limit to the amount of deviations from constant that will be recoverable, however. If the loadings vary dramatically, the optimization routine will likely find a rotation of the curves with smaller variations in the random intercept loadings. It will then select that as the final estimates. Those estimates will differ from the generating values. This however is not a bias. Simply put, PSEM has found a better rotation for the generated data than the generating rotation. This same phenomenon exists also in EFA and Alignment.

Only the simplest structures are recovered by EFA and Alignment simulations.

There is a simple way to identify if such an issue has occurred in a particular simulation study. If the PSEM estimates are different from the generating values, we can check that this reflects a simpler structure if we perform a second simulation using the PSEM estimates as the generating values. If in the second simulation, PSEM performs well, we conclude that the estimation is not biased (due to insufficient penalty or rotation) but simply that PSEM has found a model that conforms better with the desired rotation.

7.9. PSEM-ELGM Linear Growth Model

The PSEM-ELGM linear growth model can be described as follows

$$Y_t = \alpha(t) + \beta(t)I + \gamma(t)S + \varepsilon_t \quad (54)$$

$$\varepsilon_t = r\varepsilon_{t-1} + \varepsilon_t' \quad (55)$$

$$\varepsilon_t' \sim N(0, \theta_t) \quad (56)$$

$$\begin{pmatrix} I \\ S \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_I \\ \mu_S \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \quad (57)$$

$$\alpha(t) \sim ALF(0, 1) \quad (58)$$

$$DIFF(\beta(t)) \sim ALF(0, 1) \quad (59)$$

$$\gamma(1) \sim ALF(0, 1). \quad (60)$$

The parameters $\alpha(t)$, $\beta(t)$ and $\gamma(t)$ can be viewed as non-parametric functions or time specific parameters. The $\alpha(t)$ priors ensure that $\alpha(t)$ is as close to zero as possible and that the means of the factors can fit as much of the means of Y_t as possible, see Section 7.3. The Diff priors for $\beta(t)$ ensure that $\beta(t)$ is as close to constant as possible, which allows us to interpret the latent variable I as a random intercept, see Section 7.4. Finally, the prior for $\gamma(1)$ ensures that $\gamma(1)$ is as close to zero as possible. This allows us to interpret S as a growth slope factor, see Section 7.6. The last two prior specifications are responsible for curve rotation, while the first prior specification is concerned only with the mean structure.

We illustrate this model with a simulation study using $T = 8$ and $N = 1000$ observations. The Mplus input file for this study is given in SM Figure 18. We generate the data using the following setup. Two of the $\alpha(t)$ parameters are not zero, i.e., for two of the time points, the ELGM curves do not fit the mean of the observed variable. The parameter $\beta(t)$ is not a constant function but the deviations from a constant are not large. The parameter $\gamma(t)$ does not represent a linear function but the deviation is not large. The results of the simulation study are given in Table 16. The bias of the parameter estimates is minimal and the coverage is near the nominal level. The average chi-square test of fit for this study is 12.7 and with 12 degrees of freedom this yields a 7% rejection rate which is also near the nominal level. In this simulation study we used a stronger convergence criterion and a larger number for the maximum number of iterations. This is needed because the optimization is more complex than the typical SEM estimation.

Table 16. PSEM-ELGM linear growth.

Parameter	True Value	Abs. Bias	Coverage
$\alpha(1)$	0	.01	.99
$\alpha(2)$	.2	.01	.97
$\alpha(6)$	.4	.02	.94
$\beta(1)$	1	.02	.94
$\beta(5)$	.94	.02	.94
$\beta(8)$	1.06	.03	.88
$\gamma(1)$	0	.02	.99
$\gamma(5)$	.48	.04	.91
$\gamma(8)$	.86	.04	.93
r	.3	.00	.89
ρ	.3	.00	.95
μ_I	1	.01	.92
μ_S	.5	.00	.99

In this simulation study we noticed that there are some discrepancies between the standard errors and the standard deviation of the parameters across the replications for some parameters. Asymptotically, the ratio between these two quantities should be near 1 and deviations from that are generally associated with bias in the standard error estimation. In this simulation study, if the same size is increased to $N = 2000$, the discrepancy disappears. Thus, this issue is a finite sample size phenomenon. Furthermore, we notice that the discrepancy between the standard error and the standard deviation did not result in problems with the confidence interval coverage. This means that only individual replications in the simulation study had experienced problems and this is not a systematic underestimation of the standard errors. One way to resolve such problems is to increase the TOLERANCE option to 0.01 from the default value of 0.001. Increasing this option makes the computation more robust and less likely to experience issues with individual replications. Larger values of the TOLERANCE option tend to reduce the bias in the standard error estimates but also tend to increase the bias in the point estimates. Another resolution to this issue is to use bootstrap standard errors instead of the asymptotic based sandwich type standard errors, which are more susceptible to finite sample size anomalies. Yet a different resolution comes from the observation that in this simulation study, the average prior proportion is 0.61 which is too high. The proportion can be improved by increasing the prior variance using ALF(0,10) priors. This improves the prior proportion as well as the ratio between the standard errors and the standard deviations. It should be noted here that the DIFF prior tends to have more weight than non-DIFF direct priors. This is because there are many more terms involved in the DIFF prior. A DIFF prior with T parameters has $T(T-1)/2$ priors, see Equation (15), instead of T priors. Increasing only the DIFF prior variance will be optimal in many situations as this will compensate for the outsized influence of that prior specification.

7.10. Empirical example

In this section we illustrate the PSEM-ELGM methodology with an empirical example. We use the Berkeley height data discussed in Grimm et al. (2013). For $N = 127$ subjects, height is recorded in centimeters annually from age 3 to age

17, except for age 14, for a total of $T = 14$ observations. These data are monotonically increasing which is likely to provide a modeling challenge that is different from the typical behavioral data where the main predictor curve is a random intercept. We begin the analysis by estimating the ELGM null model with auto-correlation for $m = 1, 2, 3$, as illustrated in the middle part of Figure 17. At $m = 1$, we observe that the main predictor line is not a random intercept but is closer to an increasing straight line. That is, $\lambda_1(t)$ resembles a linear function. This is not surprising, given the monotonicity of the data. Such a model can also be interpreted as a linear growth model where the random intercept and the random slope are perfectly correlated. For $m = 2$, CFI = 0.97, i.e., a two curve model appears to extract most of the information from these data. For $m = 3$, negative residual variances are found and we conclude that at the sample size of $N = 127$, it will not be possible to extract a third curve reliably. Thus we focus on rotating the two curve model, i.e., we estimate the following model

$$Y_t = \alpha(t) + \lambda_1(t)f_1 + \lambda_2(t)f_2 + \varepsilon_t \tag{61}$$

$$\varepsilon_t = r\varepsilon_{t-1} + \varepsilon_t'. \tag{62}$$

$$\varepsilon_t' \sim N(0, \theta_t) \tag{63}$$

$$\begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \tag{64}$$

$$\alpha(t) \sim ALF(0,1) \tag{65}$$

$$DIFF(\lambda_1(t) - \lambda_1(t-1)) \sim ALF(0,1) \tag{66}$$

$$\lambda_2(1) \sim ALF(0,1). \tag{67}$$

The only difference between this model and the linear growth PSEM-ELGM is that the main predictor line $\lambda_1(t)$ is not a constant line but an increasing straight line. The DIFF prior for this curve loadings is not on the loading themselves but on the differences between consecutive loadings (first derivatives) as discussed in Section 7.4. The input file for this estimation is given in SM Figure 19. We use MODEL CONSTRAINT to form the differences between consecutive loadings needed for the DIFF priors. The results of this analysis are given in Figure 3 and Table 17. Figure 3 shows $\lambda_1(t)$ and $\lambda_2(t)$ curve estimates. The main predictor line in this model remains a straight line. The second curve appears to be related to teenage years growth spurs or puberty onset/delays and it has an usual shape. The estimated curve for each individual is simply a weighted sum of these two curves where the weights are the factors f_1 and f_2. The distribution of f_1 and f_2 is given in Table 17. What is important to note here is that the weight of the first curve, i.e., the straight line, is about 20 times the weight of the second curve. This conclusion is derived from the estimated range for f_1 and f_2. The weight of the second factor is equally likely to be positive or negative. Only two of the time specific intercepts parameters $\alpha(t)$ are significant and are given in Table 17. These parameters must also be added to the estimated individual curves, as in Equation (61).

Next we illustrate the performance of the PSEM methodology with a simulation study that resembles the height data. That is, the main predictor curve is a straight line while the second curve is a piecewise zero and parabolic.
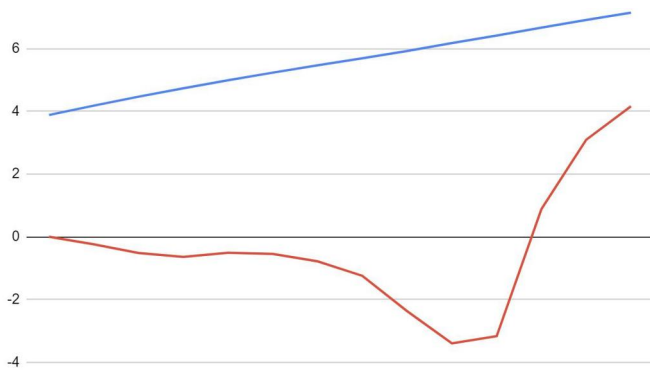
Figure 3. The two predictor curves for the height model.

Table 17. Height empirical example.

Parameter	Estimate	Stand. Error
r	.82	.03
ρ	−.01	.08
μ_1	24.70	1.77
μ_2	−.19	.10
$\alpha(12)$	3.50	.23
$\alpha(14)$	−4.35	.50

Table 18. PSEM-ELGM: Mimicking the height data.

Parameter	True Value	Abs. Bias	Coverage
$\lambda_1(1)$	1	.00	.97
$\lambda_1(5)$	1.4	.00	.96
$\lambda_1(10)$	1.9	.00	.98
$\lambda_2(5)$	0	.00	.96
$\lambda_2(6)$	.4	.00	.9
$\lambda_2(7)$	.8	.00	.95
$\lambda_2(8)$	1.2	.02	.96
$\lambda_2(9)$	.8	.02	.95
$\lambda_2(10)$	.4	.01	.96
r	.5	.00	.96
ρ	.3	.01	.97
μ_1	1	.01	.95
μ_2	.5	.00	.99
$\alpha(10)$	−.2	.01	.98

We use $T = 10$ and $N = 1000$ for this simulation study. The Mplus input file is given in SM Figure 20. One time-specific intercept is included in this model at time point 10. We use the same rotation technique as in the empirical example. The results of the simulation study for some of the parameters are given in Table 18. The bias in the parameter estimates is minimal and the coverage is near the nominal levels. The average chi-square test of fit in this simulation study is 25.6. With 25 degrees of freedom, this yields an 8% rejection rate which is near the nominal level.

In this model, we do not have a random intercept. Therefore, using the first loading of the second curve to be aimed at 0 does not have the same slope meaning as in the linear growth example from the previous section. Nevertheless, we can interpret the second curve as a sort of a deviation curve from the main line, which develops over time, i.e., it can be interpreted similar to the random slope in the linear growth model. It is of interest to know what happens if we generate data with non-zero values for $\lambda_2(1)$. For a large range of values, the solution becomes rotated to where $\lambda_2(1)$ is zero. The rotation still prefers such outcomes as the simplest solutions that minimize the penalty function.

The meaning of the simplest solution is as discussed in Section 7.8. However, not all solutions are rotated to $\lambda_2(1) = 0$. If we generate data with $\lambda_2(1)$ further away from 0, the simplest solution can have a non-zero $\lambda_2(1)$.

8. Practical Aspects of PSEM

PSEM models are less parsimonious than their SEM model counterparts. In some situations, it may be desirable and possible to convert a PSEM model to a standard SEM model and regain some parsimony. Suppose that a set of parameters in a PSEM model has ALF(0,1) priors. As an example consider the PSEM-LGCM model. The PSEM estimation determines which of these parameters are significant. We can then convert the model to a standard SEM model by removing the priors, fixing the non-significant parameters to 0 and estimating the significant parameters as free and unconstrained. We call this model the PSEM followup model. In most situations, the PSEM followup model will be identified, will likely yield data fit comparable to the PSEM model, and will be more parsimonious than the PSEM model. If the PSEM model is complex, however, the data fit of the PSEM followup model may not hold up very well as removing many insignificant parameters may still be detrimental to the data fit. If a PSEM followup model is difficult to construct, it would be necessary and acceptable to retain the PSEM model as the final model.

The main goal of the PSEM modeling framework is to extract more information from the data. It is not meant to replace approximate fit measures with exact fit measures. PSEM can be used with both. A PSEM model sometimes adds many new parameters to a standard SEM model. The goal of these new parameters is not just to reflect the data better but also to ensure that the SEM model parameters are estimated more precisely. We are obtaining a better fitting model by adding more parameters and one naturally wonders if some sort of overfitting occurs. It is easy to understand this by keeping in mind that EFA and MGA are PSEM models. To the extent that EFA is overfitting the data by freeing all loadings so is PSEM. This parallel between EFA/MGA and PSEM should be kept in mind when trying to understand PSEM. Perhaps, PSEM should be viewed as an exploratory tool. However, we prefer the interpretation that comes from MGA, where we can obtain better group model comparisons by freeing ourselves from the stringent assumptions of scalar invariance.

The PSEM optimization is more complex than the standard SEM optimization and convergence problems may arise. Usually such problems can be resolved with the various Mplus options designed for such issues: increasing the maximum number of iterations to 10000 or more, decreasing or increasing the convergence criterion, using random starting values, using as starting values the estimates of a simpler model. In addition, decreasing or increasing the penalty weight and the small ε used in the ALF, LASSO and Geomin priors can also be helpful in resolving convergence problems.

9. Conclusion

In this paper we showed how the PML estimator and the PSEM framework can be used to estimate new models that can not be estimated with existing estimation techniques. Adding parameter priors to the maximum likelihood or weighted least squares is not a new idea. It is a popular data mining technique and is often used to emulate Bayesian estimation. The PSEM methodology, however, is quite different. It is founded on the principles of EFA and multiple group alignment. The structure of PSEM is based on an unidentifiable model and a corresponding null model. The connection between the two models is defined by the penalty function. The priors of the parameters do not reflect prior knowledge but are determined by the nature of the model that we want to estimate.

We also illustrated the advantages of the ALF prior as compared to the normal prior. While the ALF penalty function is native to the MGA method, we showed that the ALF prior has advantages everywhere as it more clearly separates zero and non-zero parameters.

PSEM simplifies the technical details of the estimation because of the simple observation that when the weight of the penalty converges to zero, the estimates converge to the solution of the two stage conditional optimization where the first stage optimizes the likelihood and the second stage optimizes the objective/rotation/alignment function. This allows us to estimate new models without having to develop methodology that is specific to each problem. A clear example of that is EFA with rotation that is neither oblique or orthogonal.

PSEM is not intended to replace EFA, MGA, ESEM, AESEM, and BSEM methodologies. These methods, as implemented in Mplus, provide many additional benefits that are not available and implemented with PSEM. They are also more straightforward to use. The null model is well established, estimation is customized to guarantee success, and no monitoring of the penalty optimization is needed because with these established methods the weight can be considered infinitely close to 0. Furthermore, Mplus input files are simpler with the established methodologies because the prior is automatically constructed behind the scenes with minimal specifications. Inevitably, however, modeling efforts often reach the limits of the established methods and the frameworks can no longer accommodate specific modeling needs. The PSEM framework provides a solution for these situations.

The PSEM framework can also be infinitely customized because an arbitrary penalty function can be added to a model. Discovering new penalty functions that correspond to particular modeling needs or concepts may become the most prized discoveries. In this article, we did not provide new penalty functions. All priors and penalties we used are well known already. Clearly, there are plenty of new models that can now be estimated simply by using existing penalty designs. Perhaps a new penalty design may never be found or be needed. The regularized modeling literature is abundant with variations on the LASSO univariate priors that incrementally improve the methodology. We are less intrigued by the functional form of the penalty (ridge/lasso/ALF) but by the possibility that a new multivariate prior can be connected to a practical modeling concept, similar to how the loading matrix can be rotated by a Geomin penalty or to how the DIFF prior can be used to model invariance as well as linear and quadratic shapes in the parameters.

The PSEM method is implemented in Mplus currently for the most basic SEM model with continuous and categorical variables. The list of new models that can now be accessed with just this implementation is bound to grow. For example, growth modeling for EFA factors in longitudinal studies was not illustrated here but can now also be tackled with this method. PSEM extensions to various other modeling frameworks such as mixture models and multilevel models will open new modeling possibilities. Two such examples are mixtures of ESEM models and HLM models for EFA factors with random intercepts and slopes. The PSEM approach offers clear and straightforward access to such models, while traditional methodologies do not.

We want to conclude this article with another avenue for potential further development which arises in PSEM empirical applications. As is described here, PSEM uncompromisingly favors the data fit. In some empirical examples, however, we have noticed that a small sacrifice in the data fit can be converted to big gains in the penalty. For example, a small drop in EFA data fit can be converted to huge gains in EFA simplicity. Similarly, a small drop in data fit for multiple group alignment can be converted to a much more invariant solution. If the data fit drop is small enough as to not be a reason to reject the model then we clearly have missed out on an opportunity to discover a great model. In the PSEM framework, the trade off between the data fit and the penalty fit is controlled by the penalty weight / prior variance. The models that we are describing here lie just to the left of the PSEM bubble in Figure 2. These models can be explored in the PSEM framework by lowering the prior variance below its PSEM range, to the point where the model moves into the regularized SEM bubble and the prior variance is somewhat subjectively specified. Of course, these PSEM-RegSEM models can be justified only if a substantial model simplification (measured by an unweighted penalty reduction) is gained by the data fit sacrifice. At its extreme, we can ask the PSEM framework to return the simplest EFA structure not rejected by the data. Similarly, we can ask the PSEM framework to return the most invariant alignment model not rejected by the data. Naturally, these models are guaranteed to have a p-value of 0.05, which challenges fundamental statistical frequentist principles. Nevertheless, these PSEM-RegSEM models should not be ignored as they can provide a true value in practical settings: well fitting models that are also much simpler to interpret. They are simply not the maximum-likelihood models.

References

Arbuckle, J., & Friendly, M. L. (1977). On rotating to smooth functions. *Psychometrika, 42*, 127–140. https://doi.org/10.1007/BF02293749

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397–438. https://doi.org/10.1080/10705510903008204

Asparouhov, T., & Muthén, B. (2012). *Comparison of computational methods for high dimensional item factor analysis*. http://www.statmodel.com/download/HighDimension.pdf

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

Asparouhov, T., & Muthén, B. (2023a). Residual structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 30, 1–31. https://doi.org/10.1080/10705511.2022.2074422

Asparouhov, T., & Muthén, B. (2023b). Multiple group alignment for exploratory and structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 30, 169–191. https://doi.org/10.1080/10705511.2022.2127100

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer. *Journal of Management*, 41, 1561–1577. https://doi.org/10.1177/0149206315591075

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

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111–150. https://doi.org/10.1207/S15327906MBR3601_05

Ferrando, P. J., Hernandez-Dorado, A., & Lorenzo-Seva, U. (2022). Detecting correlated residuals in exploratory factor analysis: New proposals and a comparison of procedures. *Structural Equation Modeling: A Multidisciplinary Journal*, 29, 630–638. https://doi.org/10.1080/10705511.2021.2004543

Geminiani, E., Marra, G., & Moustaki, I. (2021). Single- and multiple-group penalized factor analysis: A trust-region algorithm approach with integrated automatic multiple tuning parameter selection. *Psychometrika*, 86, 65–95. https://doi.org/10.1007/s11336-021-09751-8

Grimm, K. J., Steele, J. S., Ram, N., & Nesselroade, J. R. (2013). Exploratory latent growth models in the structural equation modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal*, 20, 568–591. https://doi.org/10.1080/10705511.2013.824775

Hirose, K., & Yamamoto, M. (2014). Estimation of an oblique structure via penalized likelihood factor analysis. *Computational Statistics and Data Analysis*, 79, 120–132. https://doi.org/10.1016/j.csda.2014.05.011

Hirose, K., & Yamamoto, M. (2015). Sparse estimation via nonconcave penalized likelihood in factor analysis model. *Statistics and Computing*, 25, 863–875. https://doi.org/10.1007/s11222-014-9458-0

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 55–67. https://doi.org/10.1080/00401706.1970.10488634

Huang, P. (2018). A penalized likelihood method for multi-group structural equation modelling. *The British Journal of Mathematical and Statistical Psychology*, 71, 499–522. https://doi.org/10.1111/bmsp.12130

Huang, P., Chen, H., & Weng, L. (2017). A penalized likelihood method for structural equation modeling. *Psychometrika*, 82, 329–354. https://doi.org/10.1007/s11336-017-9566-9

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

Jennrich, R. (2006). Rotation to simple loadings using component loss functions: The oblique case. *Psychometrika*, 71, 173–191. https://doi.org/10.1007/s11336-003-1136-B

Lüdtke, O., Ulitzsch, E., & Robitzsch, A. (2021). A comparison of penalized maximum likelihood estimation and Markov Chain Monte Carlo techniques for estimating confirmatory factor analysis models with small sample sizes. *Frontiers in Psychology*, 12, 615162. https://doi.org/10.3389/fpsyg.2021.615162

Magis, D., Tuerlinckx, F., & De Boeck, P. (2015). Detection of differential item functioning using the lasso approach. *Journal of Educational and Behavioral Statistics*, 40, 111–135. https://doi.org/10.3102/1076998614559747

Morin, A. J. S., & Asparouhov, T. (2018). *Estimation of a hierarchical Exploratory Structural Equation Model (ESEM) using ESEM-within-CFA*. Substantive Methodological Synergy Research Laboratory. https://www.statmodel.com/download/Webnote%20-%20Hierarchical%20Exploratory%20Structural%20Equation%20Model.pdf

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, B., & Satorra, A. (1995). Technical aspects of Muthen's Liscomp approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, 60, 489–503. https://doi.org/10.1007/BF02294325

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén. 1998–2017.

Scharf, F., & Nestler, S. (2019). Should regularization replace simple structure rotation in exploratory factor analysis? *Structural Equation Modeling: A Multidisciplinary Journal*, 26, 576–590. https://doi.org/10.1080/10705511.2018.1558060

Sterba, S. K. (2014). Fitting nonlinear latent growth curve models with individually-varying time points. *Structural Equation Modeling*, 21, 630–647. https://doi.org/10.1080/10705511.2014.919828

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

Trendafilov, N. T., Fontanella, S., & Adachi, K. (2017). Sparse exploratory factor analysis. *Psychometrika*, 82, 778–794. https://doi.org/10.1007/s11336-017-9575-8

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 23, 19–23. https://doi.org/10.1007/BF02288975

Tucker, L. R. (1966). Learning theory and multivariate experiment: Illustration by determination of generalized learning curves. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 476–501). Rand McNally.

Tutz, G., & Schauberger, G. (2015). A penalty approach to differential item functioning in Rasch models. *Psychometrika*, 80, 21–43. https://doi.org/10.1007/s11336-013-9377-6

Zou, H., Hastie, T., & Tibshirani, R. (2007). On the "degrees of freedom" of the lasso. *The Annals of Statistics*, 35, 2173–2192. https://doi.org/10.1214/009053607000000127