Penalized Structural Equation Models

Tihomir Asparouhov

Mplus

Introduction

- PSEM is similar to regularized SEM in that penalty is added to the fit function.
- PSEM is not quite regularized SEM. PSEM is based on 2 models with the same likelihood: null and penalized models. Penalized model is always unidentified if penalty is removed, i.e., the penalty main function is to identify the PSEM extension of the null model
- The penalty function is specific to the type of model we estimate. If we estimate an EFA model we use Geomin penalty
- Penalty is interpreted and constructed as Prior for model parameter
- Underlying the existence of PSEM is the vision that: Penalty=Prior=Rotation/Alignment

Introduction

- PSEM is meant to be a generalization for EFA/ESEM/MG-Alignment/Logitudinal Alignment/ASEM/AESEM
- All of these methods are based on conditional optimization.
- The log-likelihood or the data fit function is optimized with respect to all parameters as a first step
- The subspace of all model parameters that maximizes the data likelihood is determined. Since the model is overparameterized this is a multidimensional space
- In that subspace, a secondary function is optimized such as rotation criterion and/or alignment loss function, to determine the model closest to a desired model such as simple loading structure or scalar invariance

Introduction

- Conditional optimization is quite complex because the first stage optimal subspace is complex. Standard errors are also computationally complex and are based on implicitly solving non-linear equation.
- Longitudinal EFA and AESEM are even more complex because these use two objective functions: alignment and rotation in addition to the data fit. This implies that the conditional optimization has 3 stages and the order of the stages is important
- PSEM mimics and approximates conditional optimization by adding all objective functions together and using small weight to ensure that data fit is optimized completely before the penalty is optimized
- However, PSEM is a one step optimization. Simpler to work with and can be used in generalized settings: it can work with any penalty (rotation, alignment or any other)

Introduction

- Why did we need PSEM? We wanted to align the intercepts in a growth model (as in longitudinal alignment). Such a model doesn't even have a factor analysis measurement model. So the existing methods can't be used. Instead of developing a new conditional optimization for every model, we embraced PSEM as a simpler alternative, which applies to other models as well
- Subsequently the connections between PSEM and BSEM, ESEM, Alignment, and RegSEM were made
- Subsequently we realized that a large number of unsolved SEM problems can be solved with PSEM

- We have been using PSEM for the last 9 months and we keep discovering new things with it. Here is the list of findings from the last month
  - Some EFA/ESEM models don't need a rotation criterion - the model can estimate all loadings without rotation
  - Rotations are not always equivalent in terms of log-likelihood
  - For some models, the best rotations can be estimated from the data because maximizing the data fit determines the best rotation
  - In Longitudinal-CFA/EFA, time-invariant correlated uniqueness can be identified even if it is not identified at each individual time point
- PSEM allows us to manually estimate and customize EFA, measurement model alignment across groups and time, structural alignment of model parameters, tuckerization of curves (exploratory growth curve modeling), etc.

Introduction

- The fundamental equation of PSEM

$$(Fit\ function) + w * Penalty$$

$$Penalty = -log(Prior)$$

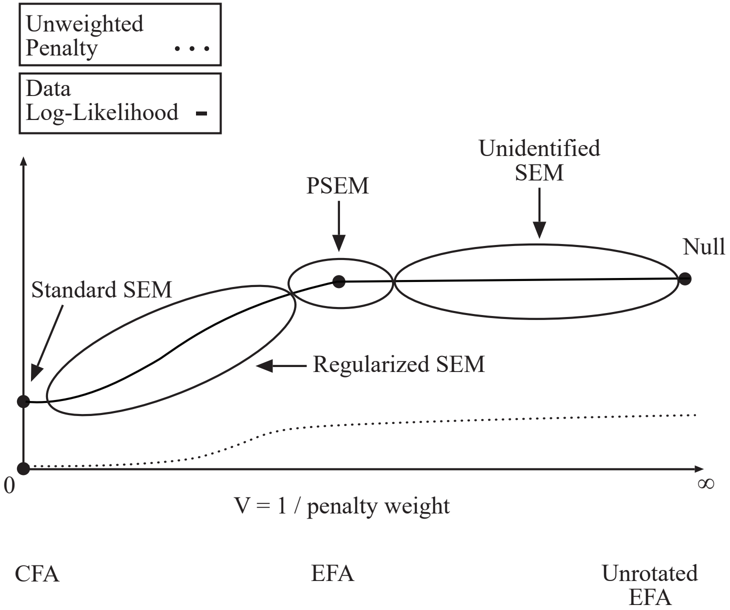
$$w = 1/(Prior\ variance)$$

- w is the penalty weight
- PSEM is based on $w \approx 0$, however, 0 or very tiny weight is not feasible due to limits in numerical precision

Introduction

- In addition to the null and the penalized model, there is also a third model in the picture which we refer to as the standard SEM model. This is when the penalty weight is infinity. The penalty is held strictly at the minimum before the data is fitted.
- This is where approximately equal parameters are equal and approximately zero parameters are 0. This is why the model is a standard SEM
- The penalty weight provides a continuum between the SEM and the null model. PSEM is on that continuum. RegSEM is also on that continuum.
- Example of model ordering SEM/PSEM/Null: CFA/EFA/Unrotated EFA
- Example of model ordering SEM/PSEM/Null: Scalar Invariance/Multiple Group Alignment/ Configural Invariance

Graphic definition of PSEM



PSEM - Theory

- The PSEM theory comes from the fact that the PSEM estimation is an approximation for conditional optimization (data fit first, then penalty) like EFA
- PSEM is equivalent to ML estimation with parameter constraints. The parameter constraints are hidden - they are the derivatives of penalty and are generally not intuitive or very explicit
- Standard errors are always MLR - sandwich based because the penalized likelihood derivatives are treated as estimating equations. Bootstrap SE can be used.
- Theory guarantees consistent point estimates and standard errors for large sample as long as PSEM is applied correctly

PSEM - Theory

- A key thing with PSEM is estimating the number of free parameters. This is not a trivial task and is determined numerically as the MLF rank. Something similar is used with regularized models but not exactly the same.
- Like in EFA - a zero loading is still a parameter, not quite how LASSO regression/regularization is counting parameters
- Counting the number of model parameters goes along the axes of PSEM and Null models. The goal is to match DF for PSEM and NULL. The NULL model is explicit only for simpler PSEM
- For complex PSEM models NULL model formulation and estimation might not be feasible. Must rely/trust the algorithm
- Simulation studies provide an easy check for correct performance. Average chi-2 must match DF.

PSEM - Overview

There are currently 6 areas where PSEM can be applied

- Structural alignment: PSEM-SA
- PSEM-ESEM for structural EFA not supported by ESEM
- ELGM: exploratory latent growth models
- Alignment: Multiple group and longitudinal invariance alignment not supported by ASEM and AESEM
- PSEM-RegSEM: models where we sacrifice some of the data fit in order to fulfill penalty demand
- Regularized SEM: PSEM models where removing the penalty is an identified model

PSEM Overview

- **PSEM-SA**
- PSEM-ESEM
- PSEM-ELGM
- PSEM-Alignment
- PSEM-RegSEM
- Regularized SEM

PSEM - SA

- PSEM - SA can be used as a substitute for BSEM (small variance prior Bayes SEM)
- An imperfect structural model or a structural model that is rejected is augmented with additional parameters which the structural model implies are zero (or not equal). PSEM is looking for a new structural model that is the best approximation for the desired structural model with minimal number of added parameters (given the use of ALF or LASSO penalty)
- Parameters are aligned to fit as best as possible a proposed SEM model. Alignment argument is derived not from invariance of measurement model as in MGA but from the structure of the model

PSEM-SA advantages over BSEM

- Based on ML and WLSMV which have more diagnostics and inference utilities
- PSEM-SA is faster than BSEM
- PSEM-SA is less sensitive to prior variance specification
- PSEM-SA requires fewer number of runs than BSEM
- PSEM-SA is more stable
- PSEM-SA can use random starting values and deal with multiple solutions
- PSEM-SA has access to the ALF priors which deliver parsimonious models
- PSEM-SA has easier specification techniques (avoids IW priors)
- If prior variance is too large the model is unidentified: PSEM uses information matrix condition number diagnostics. BSEM uses inefficient poor convergence performance
- PSEM-SA has the null model log-likelihood value as a great reference point which can show if prior variance is too small/big. Prior proportion ratio in PSEM is similarly helpful

Examples of PSEM-SA

- PSEM Growth models with time specific intercept parameters at all time points in addition to means for random intercepts and slopes.
- CFA/SEM/EFA/ESEM with full residual covariance for the indicators to discover a minimal number of correlated uniqueness needed to support the model
- In MIMIC models: adding all direct effect from covariates to indicators in addition to the covariate effects on the factors
- Cross loadings detection - adding all cross loadings - better than standard Target EFA

Examples of PSEM-SA

- PSEM Growth model

$$Y_{it} = \alpha(t) + I_i + S_i \cdot t + \varepsilon_{it} \tag{1}$$

$$\varepsilon_{it} = r\varepsilon_{i,t-1} + \varepsilon'_{it} \tag{2}$$

$$\varepsilon'_{it} \sim N(0, \theta_t) \tag{3}$$

$$\begin{pmatrix} I_i \\ S_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_I \\ \mu_S \end{pmatrix}, \begin{pmatrix} v_I & \rho \\ \rho & v_S \end{pmatrix} \right) \tag{4}$$

$$\alpha(t) \sim LASSO(0, 1) \tag{5}$$

PSEM Growth model simulation

```
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);
```

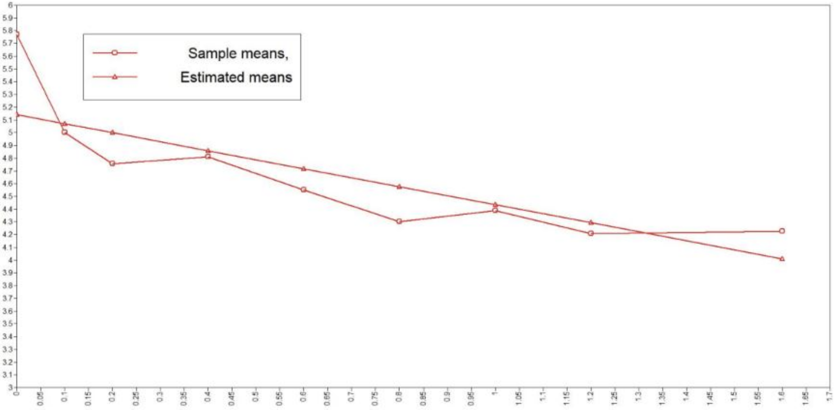
PSEM Growth model simulation results

	Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
I WITH							
S	0.300	0.2956	0.0489	0.0467	0.0024	0.940	1.000
Means							
I	0.400	0.3973	0.0576	0.0549	0.0033	0.930	1.000
S	0.100	0.1105	0.0446	0.0493	0.0021	0.960	0.610
Intercepts							
Y1	0.000	0.0005	0.0364	0.0390	0.0013	1.000	0.000
Y2	0.000	-0.0015	0.0416	0.0394	0.0017	0.970	0.030
Y3	0.000	0.0011	0.0415	0.0397	0.0017	0.990	0.010
Y4	0.000	0.0064	0.0446	0.0426	0.0020	0.960	0.040
Y5	-0.300	-0.3021	0.0474	0.0515	0.0022	0.960	1.000
Y6	0.200	0.1827	0.0604	0.0540	0.0039	0.930	0.920
Y7	0.000	-0.0102	0.0456	0.0387	0.0022	0.970	0.030
Y8	0.000	-0.0119	0.0316	0.0251	0.0011	0.940	0.060
Variances							
I	0.900	0.8905	0.0718	0.0757	0.0052	0.960	1.000
S	0.400	0.3995	0.0603	0.0574	0.0036	0.920	1.000
Residual Variances							
Y1	1.000	0.9899	0.0684	0.0728	0.0047	0.980	1.000

PSEM Growth model: Stress in the COMBINE Study (T = 9)

- Data from COMBINE, a 16-week, multisite randomized double-blind clinical trial comparing treatments of alcohol dependence (Anton et al., 2006, JAMA)
- Measurement occasions: Baseline, week 1, week 2, week 4, week 6, week 8, week 10, week 12, week 16 and week 52 follow-up (first 9 timepoints used here, week 0-week 16)
- Consider 4 models:
  - The SEM model is the standard growth model
  - The PSEM growth model from previous slide
  - The Null model where $\mu_I = \mu_S = 0$ and $\alpha(t)$ is not constrained by LASSO
  - The PSEM-follow up model where only significant $\alpha(t)$ are retained without LASSO.

Standard growth model fit



PSEM Growth model

```
Analysis: estimator = mlr;

Model:
i s | y0@0 y1@.1 y2@.2 y3@.4 y4@.6 y5@.8 y6@1 y7@1.2 y8@1.6;
y1^-y8^ pon y0^-y7^ (r);
[y0-y8] (m0-m8);

Model prior: m0-m8~LASSO(0,1);
```

PSEM Growth model results

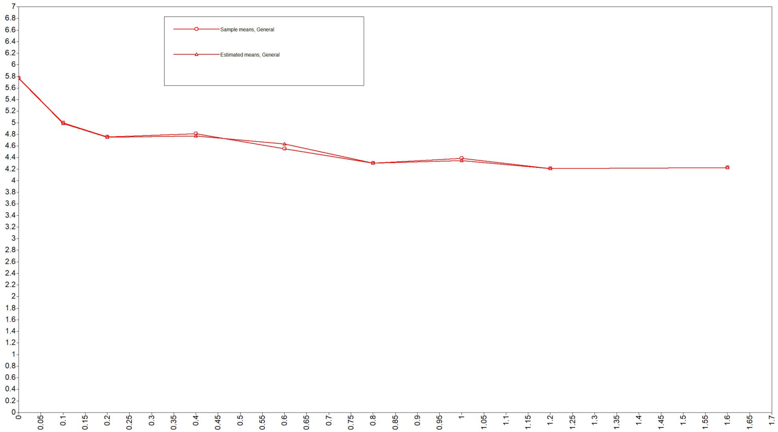
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Means				
I	5.040	0.083	60.827	0.000
S	-0.668	0.078	-8.553	0.000
Intercepts				
Y0	0.723	0.071	10.163	0.000
Y1	0.021	0.027	0.774	0.439
Y2	-0.151	0.056	-2.694	0.007
Y3	0.022	0.041	0.536	0.592
Y4	-0.091	0.063	-1.439	0.150
Y5	-0.207	0.065	-3.208	0.001
Y6	0.011	0.036	0.310	0.757
Y7	-0.020	0.039	-0.522	0.602
Y8	0.256	0.094	2.728	0.006

PSEM Growth model BIC results

Table: PSEM growth model

Model	BIC
Standard LG	46263
Null	46123
PSEM LG	46123
PSEM LG follow-up	46104

PSEM follow up model fit



5 out of 9 sample means actually are on a straight line

Alignment of cross-loadings simulation study

```
MONTECARLO:
        NAMES = y1-y6;
        NOBSERVATIONS = 500;
        NREPS = 100;

MODEL POPULATION:
        f1 by y1-y3*1 y6*0.3;
        f2 by y4-y6*1;
        f1-f2@1;
        f1 with f2*.25;
        y1-y6*.5;

MODEL:
        f1 by y1-y3*1
          y4-y5*0 y6*0.3 (a1-a3);
        f2 by y4-y6*1
          y1-y3*0 (a4-a6);
        f1-f2@1;
        f1 with f2*.25;
        y1-y6*.5;

MODEL PRIORS: a1-a6~ALF(0,1);
```

Table: 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)

Alignment of cross-loadings

- ALF and LASSO are generally similar. Only for larger sample sizes 10000, ALF becomes better and LASSO yields significant biases. ALF is more aggressive in seeking parsimonious models.
- In our Alignment development LASSO $|x|$ was not in the top two links, ALF $\sqrt{|x|}$ was the best and $\sqrt[4]{|x|}$ was second best
- EFA-Target is identical to PSEM-Normal. Rotation criterion = Penalty.
- EFA-Target may be obsolete as it can not compete with PSEM-ALF.
- The main drawback of EFA-Target is that if a target is not zero, it must be removed to improve performance, becomes a multi-stage iterative procedure as in CFA-ModInd
- BSEM requires multiple runs to get good results but it can be adjusted to yield results similar to PSEM-Normal

Alignment of residual correlations in EFA

- PSEM can estimate an EFA model and LASSO-include all residual correlations

```
MODEL:
 f1-f2 by u1-u10*0 (*1);
 u1-u10 with u1-u10 (c1-c45);

MODEL PRIORS: c1-c45~LASSO(0,1);
```

Alignment of residual correlations in EFA simulation study

- 2 factors, 10 indicators, 1 cross loading, and 3 residual covariances, N=500
- ALF prior is better almost universally in PSEM, LASSO used mostly for comparative purposes.
- ALF bias is 0 at N=5000 as theory guarantees

```
LASSO PRIOR
 U1       WITH
  U2                0.500     0.4181     0.0869     0.0835     0.0142 0.850 1.000
 U4       WITH
  U8                0.300     0.2574     0.0641     0.0607     0.0059 0.850 1.000
 U9       WITH
  U10               0.400     0.3315     0.0701     0.0757     0.0096 0.870 0.990

ALF PRIOR
 U1       WITH
  U2                0.500     0.4503     0.1147     0.0846     0.0155 0.869 0.990
 U4       WITH
  U8                0.300     0.2767     0.0704     0.0694     0.0054 0.939 0.949
 U9       WITH
  U10               0.400     0.3672     0.0710     0.0864     0.0061 0.960 0.990
```

The PSEM recovery concept

- The same concept exist in EFA and Alignment, and could be the most misunderstood concept in SEM
- If you construct a model for a simulation study, will PSEM recover the parameter estimates?
- PSEM-SA-ALF recovers the ”most parsimonious” models only. If you generate data from a less parsimonious model and PSEM finds another equivalent model that is more parsimonious then it won’t be recovered. This doesn’t mean that PSEM is biased
- The concept only applies to simulation studies. By definition PSEM real data analysis yields the most parsimonious model
- Asymptotically this is all clear but for finite sample size things may get messier
- Example: if we add 20 residual correlations in the above EFA model - PSEM will not recover the generating parameter and will yield another model with 10 residual correlations

PSEM Overview

- PSEM-SA
- **PSEM-ESEM**
- PSEM-ELGM
- PSEM-Alignment
- PSEM-RegSEM
- Regularized SEM

PSEM-ESEM

- EFA/ESEM can be estimated with PSEM by specifying Geomin prior for a loading matrix. Penalty becomes exact replica of Geomin rotation criteria.
- Other rotations can be specified manually by constructing new parameters in model constraint and then specifying a prior for those new parameters.
- Regularized EFA models: LASSO prior is specified for every loading parameter, suggested for large P EFA models, Scharf & Nestler (2019): Should Regularization Replace Simple Structure Rotation in Exploratory Factor Analysis?, SEM Journal.
- ESEM and PSEM can be combined for some models where the rotation can be done with traditional rotation while PSEM penalty can be used for another purpose such as free residual covariances of indicators

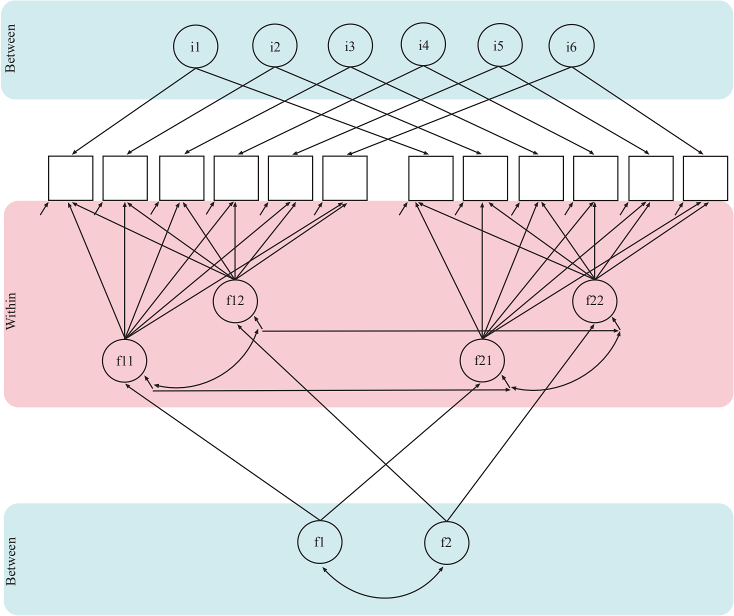
PSEM-ESEM Applications

- PSEM can be given starting values and may give better convergence rates than EFA, which estimates the more difficult unrotated model. Useful in gradually increasing the number of factors.
- Structural EFA models may not be available with ESEM. ESEM has many restrictions: not just the loadings is rotated but the entire model
- Regressions / path analysis among EFA factors
- Partial EFA invariance for multiple groups or longitudinal models
- Scalar or Metric invariant multiple group or longitudinal EFA with orthogonal rotation. ESEM yields orthogonal rotation only in the reference block.

PSEM-ESEM Applications

- Growth and various AR modeling for EFA factors
- Rotations that are neither orthogonal or oblique, i.e., factors are not independent but not completely unrestricted as in oblique rotation
  - Bi-factor EFA with 2 or more general factors
  - Second order factor analysis. There are three new models:
    - First order CFA, second order EFA
    - First order EFA, second order CFA
    - First order EFA, second order EFA

PSEM-ESEM Example: RI-CLPM for EFA factors



RI-CLPM for EFA factors

- The lagged model can be CLPM or only AR
- Without the RI, PSEM is not needed and ESEM can estimate the CLPM
- Illustration is based on Bengt's example: 6 positive affect items measure 2 EFA factor over a T=7 periods
- The model uses scalar invariance for the factors
- RI for all indicators are independent which allows us to estimate RI for the factors (one or the other)

```
model:
! random intercepts for all 6 items:
i1 by relax1-relax7@1;
i2 by satis1-satis7@1;
i3 by conf1-conf7@1;
i4 by happy1-happy7@1;
i5 by energ1-energ7@1;
i6 by excit1-excit7@1;

! auto-regressions among factor indicator residual:
relax2^-relax7^ pon relax1^-relax6^ (ar1);
satis2^-satis7^ pon satis1^-satis6^ (ar2);
conf2^-conf7^ pon conf1^-conf6^ (ar3);
happy2^-happy7^ pon happy1^-happy6^ (ar4);
energ2^-energ7^ pon energ1^-energ6^ (ar5);
excit2^-excit7^ pon excit1^-excit6^ (ar6);

! 2-factor ESEM for each of the 7 time points:
f11-f12 by relax1*1 satis1 conf1 happy1 energ1 excit1 (a1-a12);
f21-f22 by relax2*1 satis2 conf2 happy2 energ2 excit2 (a1-a12);
f31-f32 by relax3*1 satis3 conf3 happy3 energ3 excit3 (a1-a12);
f41-f42 by relax4*1 satis4 conf4 happy4 energ4 excit4 (a1-a12);
f51-f52 by relax5*1 satis5 conf5 happy5 energ5 excit5 (a1-a12);
f61-f62 by relax6*1 satis6 conf6 happy6 energ6 excit6 (a1-a12);
f71-f72 by relax7*1 satis7 conf7 happy7 energ7 excit7 (a1-a12);

! Oblique rotation
f11 f21 f31 f41 f51 f61 f71 PWITH
f12 f22 f32 f42 f52 f62 f72;
```

RI-CLPM for EFA factors

```
! Random intercept for the EFA factors
f1 by f11@1 f21@1 f31@1 f41@1 f51@1 f61@1 f71@1; f1*1;
f2 by f12@1 f22@1 f32@1 f42@1 f52@1 f62@1 f72@1; f2*1;

! CLPM for the factors
f21^-f22^ on f11^-f12^;
f31^-f32^ on f21^-f22^;
f41^-f42^ on f31^-f32^;
f51^-f52^ on f41^-f42^;
f61^-f62^ on f51^-f52^;
f71^-f72^ on f61^-f62^;

! Scalar Invariance
[relax1 satis1 conf1 happy1 energ1 excit1] (m1-m6);
[relax2 satis2 conf2 happy2 energ2 excit2] (m1-m6);
[relax3 satis3 conf3 happy3 energ3 excit3] (m1-m6);
[relax4 satis4 conf4 happy4 energ4 excit4] (m1-m6);
[relax5 satis5 conf5 happy5 energ5 excit5] (m1-m6);
[relax6 satis6 conf6 happy6 energ6 excit6] (m1-m6);
[relax7 satis7 conf7 happy7 energ7 excit7] (m1-m6);

[f11-f12@0]; [f21-f72*0];
f11-f12@1; f21-f72*1;
i1-i6 with i1-i6@0;
i1-i6 with f1-f2@0;
i1-i6 with f11-f72@0;

model prior: a1-a12~Geomin(2,1);
```

RI-CLPM for EFA factors results

- Results for this model are pretty competitive. Better BIC (13796) than the model without factor RI and with correlated RI for the indicators (13812).
- CFA for this model gives better BIC (13785) than EFA. The CFA mode doesn't need PSEM.
- Replacing CLPM with TI-AR (time invariant auto-regression) in EFA or CFA for the factors yields even better results BIC=13731
- EFA with TI-AR (13737) does not need rotation (no PSEM): native model rotation estimated from the data. CL and rotation are interconnected. Stability of factor as rotation criterion.
- Native model rotation: if you keep lowering the penalty weight and log-likelihood keeps improving and is identified, the penalty can be taken out. Loadings are identified without rotation criterion. Requires AR>0, in this case 0.5 and 0.2. Native rotation exits iff Null doesn't exist
- Another model that requires PSEM-ESEM is if you estimate EFA for the indicator intercepts I_j

Hierarchical EFA

- 20 items measure 4 EFA factors
- The 4 EFA factors measure 1 second order factor

$$Y = \nu + \Lambda_1 F + \varepsilon$$

$$F = \Lambda_2 \eta + \xi$$

- Rotation is not oblique or orthogonal since EFA factor variance covariance is structured
- Estimation can use loading starting values from the oblique EFA

```
analysis: starts=10; iter=10000;

MODEL:
 f1-f4 BY y1*0 y2-y20 (a1-a80); f1-f4@1;
 f0 by f1-f4*1; f0@1;

model prior:
a1-a80~Geomin(4,.1,.001);
```

HEFA Simulation study results N=2000

		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
F1	BY							
Y1		0.700	0.6943	0.0307	0.0305	0.0010	0.940	1.000
Y2		1.300	1.2916	0.0453	0.0466	0.0021	0.950	1.000
Y3		0.800	0.7950	0.0321	0.0327	0.0010	0.950	1.000
Y4		0.800	0.7943	0.0296	0.0326	0.0009	0.970	1.000
Y5		0.300	0.2953	0.0270	0.0272	0.0007	0.930	1.000
Y6		0.000	-0.0007	0.0210	0.0221	0.0004	0.990	0.010
....								
F2	BY							
Y1		0.000	-0.0014	0.0208	0.0221	0.0004	0.980	0.020
Y2		0.000	0.0018	0.0170	0.0162	0.0003	1.000	0.000
Y3		0.000	-0.0020	0.0209	0.0218	0.0004	0.970	0.030
Y4		0.000	-0.0025	0.0224	0.0221	0.0005	0.970	0.030
Y5		0.600	0.5969	0.0291	0.0307	0.0008	0.970	1.000
Y6		0.700	0.6961	0.0307	0.0320	0.0009	0.960	1.000
Y7		0.500	0.4992	0.0330	0.0276	0.0011	0.910	1.000
Y8		0.800	0.7959	0.0345	0.0347	0.0012	0.950	1.000
....								
F0	BY							
F1		1.400	1.4187	0.0693	0.0763	0.0051	0.970	1.000
F2		1.500	1.5289	0.0873	0.0902	0.0084	0.980	1.000
F3		1.500	1.5221	0.1017	0.1016	0.0107	0.960	1.000
F4		1.000	1.0131	0.0553	0.0609	0.0032	0.980	1.000

PSEM Overview

- PSEM-SA
- PSEM-ESEM
- **PSEM-ELGM**
- PSEM-Alignment
- PSEM-RegSEM
- Regularized SEM

- Exploratory Latent Growth Curve Model: factor rotation to growth curve shapes.
- The idea dates back to Tucker (1958, 1966) and more recently in Grimm et al. (2013) with ESEM
- It expands PSEM-LG by freeing all loadings
- PSEM-LG fits means structure perfectly, PSEM-ELGM aims at fitting the variance covariance perfectly as good as EFA

$$Y_{it} = \alpha(t) + \lambda_1(t) I_i + \lambda_2(t) S_i + \varepsilon_{it} \tag{6}$$

$$\varepsilon_{it} = r\varepsilon_{i,t-1} + \varepsilon'_{it} \tag{7}$$

$$\varepsilon'_{it} \sim N(0, \theta_t) \tag{8}$$

$$\begin{pmatrix} I_i \\ S_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_I \\ \mu_S \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \tag{9}$$

$$\alpha(t) \sim LASSO(0,1) \tag{10}$$

$$\lambda_1(t), \lambda_2(t) \sim ? \tag{11}$$

PSEM-ELGM

- Factor rotation is desired to get the most interpretable model, i.e., rotation criteria that yield growth curve shapes
- Rotations need to be custom made
- Rotations must NOT aim for simple loading structure. No Geomin
- Random intercept loads on all equally. Random developmental slope loads progressively on all.
- First step in such growth models is to establish the number of latent factors with EFA and EFA-AR
- The NULL for PSEM-ELGM is EFA and EFA-AR
- PSEM-ELGM estimated individual trajectories are a weighted sum of the three curves: $\alpha(t)$, $\lambda_1(t)$ and $\lambda_2(t)$. The weights are 1 and the factor scores for I and S

PSEM-ELGM

- Rotation criterion that seems to work universally well for a two factor ELGM

$$DIFF(\lambda_1(t)) \sim LASSO(0,1), \lambda_2(1) \sim LASSO(0,1) \tag{12}$$

- The DIFF prior/penalty consists of a sum of pairwise differences in a group of parameters. It is used to establish equality among the parameters. Similar to a multivariate normal prior with correlations .9999 used in BSEM, but for ALF/LASSO

$$DIFF(P_1 - P_N) \sim LASSO(0,1)$$

$$Penalty = \sum_{i<j} |p_i - p_j|$$

- The DIFF prior sets the first factor as a random intercept and the second as a developmental curve
- It is tempting to put more structure on the slope curve but it is not necessary as there are only 2 parameters to identify

PSEM-ELGM application

- Using the PSEM-LG example discussed earlier: Stress in the COMBINE Study (T = 9) over 16 weeks

Table: PSEM growth model

Model	LL	NP	chi-2	p-val	BIC
PSEM LG	-22982	22	122	0	46123
PSEM LG follow-up	-22983	19	127	0	46104
EFA	-22947	35	64	0	46148
EFA-AR	-22924	36	27	.06	46109
PSEM-ELGM-AR	-22924	36	27	**.06**	46109
PSEM-ELGM-AR follow-up	-22941	23	58	.002	**46049**

Both PSEM-ELGM-AR and PSEM-ELGM-AR followup are of interest depending on the preferred fit criteria. The follow-up retains significant parameters, and OUTPUT:ALIGN scrutinizes equality for the DIFF loadings: only the first loading is significantly different.

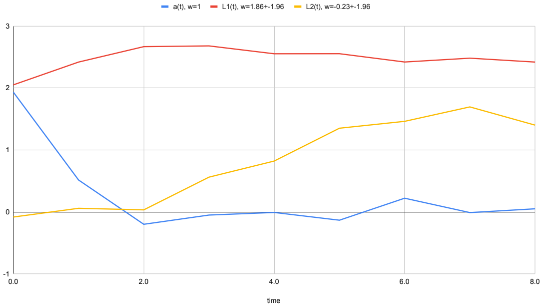
Comparing loading matrices

$$\Lambda_{EFA} = \begin{pmatrix} 2.2 & 2.4 & 2.7 & 1.8 & 1.2 & 0.3 & 0.0 & -0.3 & 0.1 \\ -0.2 & 0.1 & 0.0 & 0.9 & 1.4 & 2.3 & 2.5 & 2.9 & 2.4 \end{pmatrix}$$

$$\Lambda_{ELGM} = \begin{pmatrix} 2.1 & 2.4 & 2.7 & 2.7 & 2.6 & 2.6 & 2.4 & 2.5 & 2.4 \\ -0.1 & 0.1 & 0.0 & 0.6 & 0.8 & 1.4 & 1.5 & 1.7 & 1.4 \end{pmatrix}$$

Better interpretation for ELGM: ”baseline” and ”growth development”

The three curves of PSEM-ELGM: Stress development over 16 weeks



- Blue=$\alpha(t)$: population level adjustment, weight=1
- Red=$\lambda_1(t)$: baseline curve, weight=I=1.86+-1.96
- Yellow=$\lambda_2(t)$: growth curve, weight=S=-0.23+-1.96

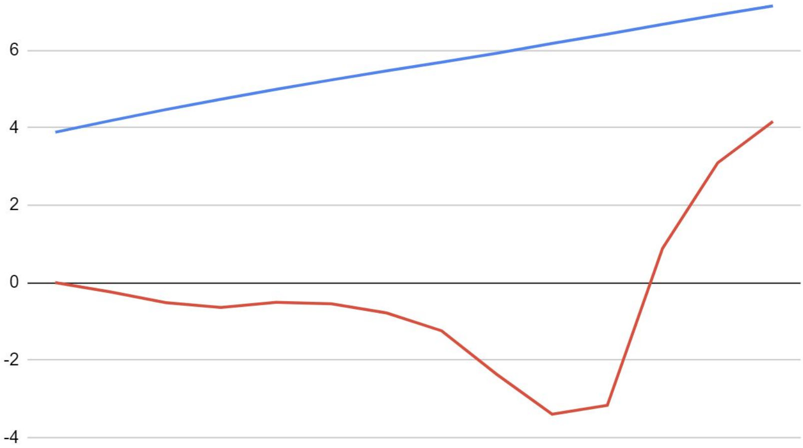
PSEM-ELGM Height in adolescents example

- Modeling height in adolescents: the "baseline" curve for height continuously increases, Asparouhov & Muthén (2023)
- One possibility is to NOT use a random intercept but instead to use a linear growth curve $\lambda t I_i$ as the baseline curve
- This means that the loadings are approximately a linear function of t
- ELGM rotation

$$DIFF(\lambda_1(t) - \lambda_1(t-1)) \sim LASSO(0,1)$$

$$\lambda_2(1) \sim LASSO(0,1)$$

PSEM-ELGM Height in adolescents example



Baseline linear growth curve and a puberty growth spurt adjustment curve. Individual height growth is a weighted sum of the two curves.

PSEM Overview

- PSEM-SA
- PSEM-ESEM
- PSEM-ELGM
- **PSEM-Alignment**
- PSEM-RegSEM
- Regularized SEM

PSEM-Alignment

- Alignment is used to establish approximate measurement invariance across time or across groups. Automatic implementation via ALIGNMENT=FIXED/FREE option(conditional optimization)
- AESEM/ASEM expands Alignment to general ESEM, SEM and longitudinal models. Asparouhov and Muthén (2022).
- PSEM-Alignment can be used when AESEM/ASEM models are not available. AESEM/ASEM models are fairly restricted.
- Using DIFF(parameters)$\sim$ALF(0,1) establishes approximate invariance for a group of parameters and yields a penalty that is nearly identical to alignment loss function.
- Huang (2018) also uses regularized SEM for Alignment

PSEM-Alignment Applications

- Growth modeling for aligned factors
- Metric alignment: alignment for loadings only
- Separately establishing metric and scalar alignment: splitting means and covariance fit
- Partial alignment models: if certain indicators have a lot of non-invariance it is better to exclude them from the alignment
- Between level predictors for aligned factors (coefficient is invariant across group/time)
- RI-AR and RI-CLPM for aligned factors

PSEM-Alignment Example

Longitudinal EFA with 4 time points, 10 indicators, 2 factors, Metric invariance, scalar alignment, RI-AR for each factor. Alignment, ESEM, RSEM combination estimated with PSEM.

```
MODEL:
    f11-f12 BY y11-y20*1 (a1-a20);
    f21-f22 BY y21-y30*1 (a1-a20);
    f31-f32 BY y31-y40*1 (a1-a20);
    f41-f42 BY y41-y50*1 (a1-a20);
    f11-f12@1; f21-f42*1;
    [f21-f42*0];
    f1 by f11@1 f21@1 f31@1 f41@1;
    f2 by f12@1 f22@1 f32@1 f42@1;
    [y11-y20] (m1-m10);
    [y21-y30] (n1-n10);
    [y31-y40] (p1-p10);
    [y41-y50] (q1-q10);
    f21^ f31^ f41^ pon f11^ f21^ f31^ (r1);
    f22^ f32^ f42^ pon f12^ f22^ f32^ (r2);

MODEL PRIORS:
    a1-a20~geomin(2,1);
    do(#,1,10) DIFF(n# m# p# q#)~ALF(0,1);
```

PSEM-Longitudinal Alignment growth modeling

Growth model for a factor with scalar alignment (no scalar invariance). Consistent estimates for time-specific invariant and non-invariant measurement parameters as well as growth factor distribution.

```
MODEL:
    f1 BY y11-y13*1 (a11-a13);
    f2 BY y21-y23*1 (a21-a23);
    f3 BY y31-y33*1 (a31-a33);
    f4 BY y41-y43*1 (a41-a43);
    f5 BY y51-y53*1 (a51-a53);
    [y11-y53] (n1-n15);
    i s | f1@0 f2@1 f3@2 f4@3 f5@4;
    [i@0]; i@1;

MODEL PRIOR:
    DIFF(a11 a21 a31 a41 a51)~ALF(0,1);
    DIFF(a12 a22 a32 a42 a52)~ALF(0,1);
    DIFF(a13 a23 a33 a43 a53)~ALF(0,1);
    DIFF(n1 n4 n7 n10 n13)~ALF(0,1);
    DIFF(n2 n5 n8 n11 n14)~ALF(0,1);
    DIFF(n3 n6 n9 n12 n15)~ALF(0,1);
```

```
F1        BY
 Y11                 1.000      0.9930     0.0499     0.0557     0.0025 0.980 1.000
 Y12                 1.000      0.9893     0.0526     0.0562     0.0029 0.960 1.000
 Y13                 1.000      0.9943     0.0496     0.0559     0.0025 0.960 1.000

F2        BY
 Y21                 1.000      0.9919     0.0516     0.0599     0.0027 0.980 1.000
 Y22                 0.500      0.5008     0.0332     0.0379     0.0011 0.990 1.000
 Y23                 1.000      0.9923     0.0531     0.0595     0.0028 0.980 1.000

Means
 I                   0.000      0.0000     0.0000     0.0000     0.0000 1.000 0.000
 S                   0.300      0.3109     0.0271     0.0281     0.0008 0.950 1.000

Variances
 I                   1.000      1.0000     0.0000     0.0000     0.0000 1.000 0.000
 S                   0.300      0.3078     0.0319     0.0341     0.0011 0.990 1.000

I         WITH
 S                  -0.200     -0.2004     0.0254     0.0261     0.0006 0.960 1.000

Intercepts
 Y11                 0.000     -0.0084     0.0498     0.0520     0.0025 0.950 0.050
 Y12                 0.000      0.0003     0.0492     0.0520     0.0024 0.950 0.050
 Y13                 0.000     -0.0008     0.0520     0.0521     0.0027 0.950 0.050
 Y41                 0.000     -0.0154     0.0603     0.0607     0.0038 0.960 0.040
 Y42                 0.500      0.4839     0.0633     0.0665     0.0042 0.950 1.000
 Y43                 0.000     -0.0135     0.0634     0.0608     0.0042 0.930 0.070
```

PSEM Overview

- PSEM-SA
- PSEM-ESEM
- PSEM-ELGM
- PSEM-Alignment
- **PSEM-RegSEM**
- Regularized SEM

PSEM-RegSEM

- PSEM is based on a duality of two models that have identical fit: null and penalized model.
- The penalized model includes parameters of interest that can not be identified without the penalty
- The penalty reflects a key important substantive concept usually and satisfying the penalty is of interest as well and not just the chi-square
- PSEM-RegSEM sacrifices some of the data fit to get a better fit for the penalty.
- Suppose that model M1 is the duality PSEM model and model M2 is a PSEM model with a higher penalty weight and worse fit than the null/M1 model but with much better satisfied penalty
- Suppose M2 and M1 are both not rejected by the chi-square test. M2 can be preferred on substantive grounds.

PSEM-RegSEM

- This puts us in the realm of broken duality: the PSEM-RegSEM model M2 no longer has the same likelihood as the Null
- It can be viewed as "Bayes" where prior variance/penalty weight is no longer a technical matter but can be chosen subjectively on substantive grounds
- Priors/Penalty are a form of adding parameter constraints to a model, i.e., PSEM-duality can be viewed as a regular SEM with penalty derivatives as the parameter constraints
- Similarly, PSEM-RegSEM can be viewed as a regular SEM with parameter inequality constraints (Penalty/Geomin < 1 or estimate EFA with no more than 5 cross-loadings even if you can't get the log-likelihood of the standard EFA)

- Example 1: putting higher Geomin weight than the duality weight, may reduce cross-loadings, give worse fit than EFA, but it may still give an acceptable chi-square or approximate fit and much fewer cross-loadings
- Example 2: putting higher alignment weight than the duality weight, may reduce differences across loadings/intercepts and thresholds, give worse fit than the configural model, but it may still give an acceptable chi-square or approximate fit and the model can be much closer to scalar invariance model than Alignment
- PSEM-RegSEM can estimate the best approximation to a proposed model with $CFI \geq 0.95$ or p-value≥ 0.05.
- Possibly automate what we do or is this a slippery slope where every model we use will have a p-value of .05?

PSEM Overview

- PSEM-SA
- PSEM-ESEM
- PSEM-ELGM
- PSEM-Alignment
- PSEM-RegSEM
- **Regularized SEM**

PSEM framework for estimating regularized SEM

- PSEM can be used for regularized SEM estimation, Jacobucci et alt. (2016) and the other way around
- Regularized SEM estimation is ML/WLS estimation of **identified** models with priors.
- There is no model duality for such applications
- Applications:
  - LASSO/ridge regressions as part of SEM
  - Dealing with general empirical unidentification due to small sample size
  - Models where priors can be obtained from a previous study
- There is a large body of statistical literature on regularized/penalized models

PSEM conclusion

- PSEM is a masterful technique that will produce many discoveries for SEM
- Hypothetical Example. Mplus support question: Can I regress only one of the EFA factors on gender. Mplus answer for the past 15 years: No, you have to regress all EFA factors on gender or we can't rotate your model. Mplus new answer: sure but you might want to use the model native rotation instead of Geomin
- PSEM has promising future also for Mixture and Multilevel models