

New Features in Mplus

Tihomir Asparouhov

Mplus

June 25, 2023

Penalized Structural Equation Models

- **Introduction to PSEM**
- Univariate Priors
- PSEM - growth model
- Multivariate priors
- PSEM-EFA example
- Arbitrary prior/penalty
- MIMIC model with direct effects
- LASSO for the residual covariances in EFA
- Exploratory Latent Growth Models
- Alignment Illustration

Penalized Structural Equation Models

- PSEM is similar to regularized SEM, penalized maximum likelihood, and LASSO regression. 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., penalty main function is to identify the penalized model extension of the null model
- The penalty function does not reflect prior knowledge but is specified by the type of model you are estimating. If you are estimating an EFA model you use Geomin penalty
- Penalty in Mplus is given with Model Prior, but penalty is not a reflection of prior knowledge
- Underlying the existence of PSEM is:
Penalty=Prior=Rotation/Alignment

Penalized Structural Equation Models

- PSEM allows us to manually do and customize EFA, measurement model alignment across groups and time, structural alignment of model parameters, tuckerization of curves (exploratory growth curve modeling), etc.
- The fundamental equation of PSEM

$$(\textit{Fit function}) + w * \textit{Penalty}$$

$$\textit{Penalty} = -\log(\textit{Prior})$$

$$w = 1/(\textit{Prior variance})$$

- 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
- We generally start with $w = 1$ and progress to decrease w to make sure that the penalized model reaches the fit of the null model. $w = 1$ yields prior=1 person on approximately standardized scale.

- Introduction to PSEM
- **Univariate Priors**
- PSEM - growth model
- Multivariate priors
- PSEM-EFA example
- Arbitrary prior/penalty
- MIMIC model with direct effects
- LASSO for the residual covariances in EFA
- Exploratory Latent Growth Models
- Alignment Illustration

- Normal: $p \sim N(0, v)$ implies p^2/v penalty
- LASSO: $p \sim \text{LASSO}(0, v)$ implies $|p|/v$ penalty
- ALF: $p \sim \text{ALF}(0, v)$ implies $\sqrt{|p|}/v$ penalty
- All 3 are typically used to imply that a parameter should be zero if possible.
- The TOLERANCE= ϵ option applies to ALF and LASSO to make the penalty smooth and is defaulted at 0.001. For LASSO $\sqrt{p^2 + \epsilon}/v$. For ALF $\sqrt[4]{p^2 + \epsilon}/v$
- If p is positive TOLERANCE can be set to zero (for both LASSO and ALF) and with LASSO one can get penalty of p/v
- The parameter v is inversely proportional to the penalty weight, but it can be interpreted as it is with normal prior, i.e., variance of the prior. If $v = 0$ the parameter becomes fixed.

PSEM Univariate Priors

- Normal prior tends to spread around model misfit and tends to show bias in estimates when some of the penalized parameters are not zero.
- LASSO and ALF are very similar and drive parameters all the way down to zero. ALF tends to be better than LASSO for large samples.
- Consider 3 equivalent (by log-likelihood) models $M1$, $M2$, $M3$ with 4 penalized parameters. Here is how different penalties pick "the best model"

Table: Penalty function values

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

- Each method picks the lowest penalty model. Only ALF picks the parsimonious model

- Introduction to PSEM
- Univariate Priors
- **PSEM - growth model**
- Multivariate priors
- PSEM-EFA example
- Arbitrary prior/penalty
- MIMIC model with direct effects
- LASSO for the residual covariances in EFA
- Exploratory Latent Growth Models
- Alignment Illustration

$$Y_t = v_t + I + S \cdot t + \varepsilon_t \quad (1)$$

$$\varepsilon_t = r\varepsilon_{t-1} + \varepsilon'_t, \quad (2)$$

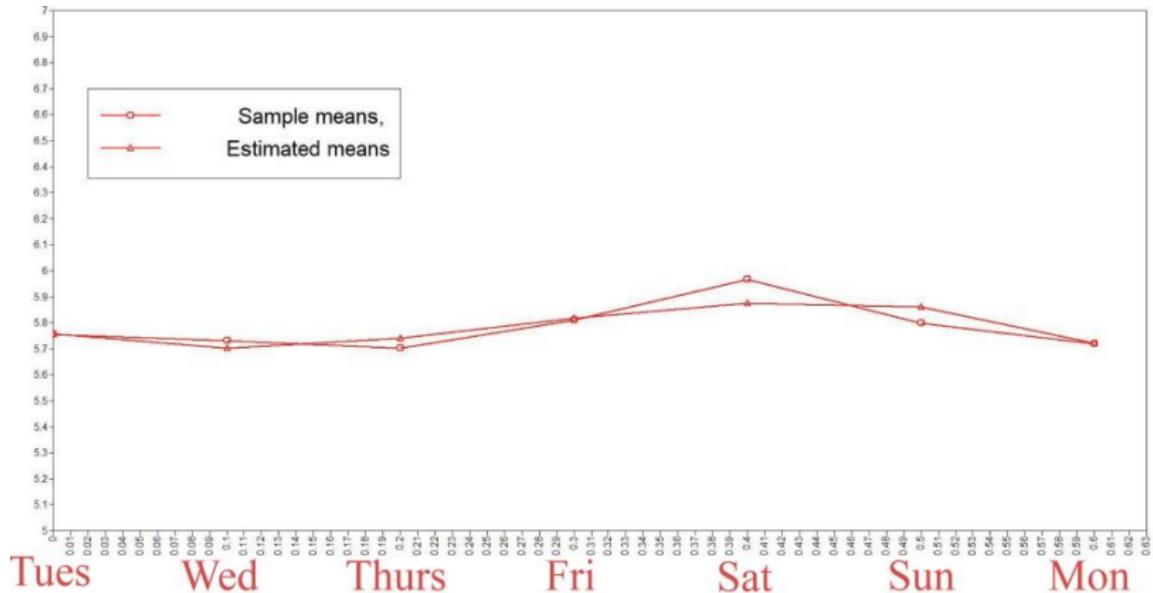
$$\varepsilon'_t \sim N(0, \sigma_t) \quad (3)$$

$$\begin{pmatrix} I \\ S \end{pmatrix} \sim N \left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \begin{pmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{pmatrix} \right) \quad (4)$$

- Standard linear growth model + time specific intercept v_t .
- Without a penalty the model is unidentified. $T+2$ mean parameters v_t , α , β and only T degrees of freedom to fit \bar{Y}_t
- We use ALF priors for v_t . The goal of the ALF prior is to minimize the number v_t parameters that are needed to avoid misfit
- This model extracts α and β while still allowing v_t to match perfectly \bar{Y}_t

- The PSEM-null model is $\alpha = \beta = 0$, i.e., this is the likelihood we must get.
- We illustrate the modeling with slide 10 data: Growth Modeling of PA: 7 Days, Tuesday - Monday

Figure: Cubic model fit



```
model:  
  i s |pa1@0 pa2@.1 pa3@.2 pa4@.3 pa5@.4 pa6@.5 pa7@.6;  
  [pa1-pa7] (m1-m7);  
  pa2^-pa7^ pon pa1^-pa6^ (ar);  
  
model prior:  
  m1-m7~ALF(0,1);
```

PSEM - growth model for PA results

Only the Saturday effect is significant. Slope mean is not significant.
Mean structure can thus be modeled with just two parameters.

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Means				
I	5.744	0.056	101.959	0.000
S	0.070	0.161	0.434	0.665
Intercepts				
PA1	0.011	0.012	0.844	0.398
PA2	-0.019	0.027	-0.722	0.470
PA3	-0.052	0.041	-1.250	0.211
PA4	0.045	0.046	0.984	0.325
PA5	0.195	0.057	3.438	0.001
PA6	0.013	0.010	1.382	0.167
PA7	-0.065	0.078	-0.832	0.405

Table: PSEM PA growth model

Model	LL	Num Param	Chi-2	p-val	BIC
Cubic	-1318	15	31	.08	2720
PSEM LG	-1313	18	21	.23	2725
PSEM LG Null	-1313	18	21	.23	2725
PSEM LG follow-up	-1318	13	29	.14	2707
PSEM RI	-1321	16	34	.02	2729
PSEM RI follow-up	-1325	11	43	.01	2710

Follow-up means we eliminate non-significant mean structure parameters and the penalty. RI is random intercept only (no slope).

- Introduction to PSEM
- Univariate Priors
- PSEM - growth model
- **Multivariate priors**
- PSEM-EFA example
- Arbitrary prior/penalty
- MIMIC model with direct effects
- LASSO for the residual covariances in EFA
- Exploratory Latent Growth Models
- Alignment Illustration

- DIFF Prior. Used to establish that parameters are equal.
 $DIFF(p_1 - p_n) \sim F(0, \nu)$, where F can be Normal/LASSO or ALF.
- The penalty then becomes.

$$\sum_{i < j} F(p_i - p_j)$$

- If F is normal the penalty becomes

$$(1/(2\nu)) \sum_{i < j} (p_i - p_j)^2$$

- It is equivalent to specifying $n(n - 1)/2$ priors for $p_i - p_j$

- Geomin Prior. Used to establish Geomin rotation for a loading matrix in EFA

$$\lambda_{11} - \lambda_{pm} \sim \text{Geomin}(m, \nu, \varepsilon)$$

implies penalty function

$$(1/\nu) \sum_i \left(\prod_j (\lambda_{ij}^2 + \varepsilon) \right)^{(1/m)}$$

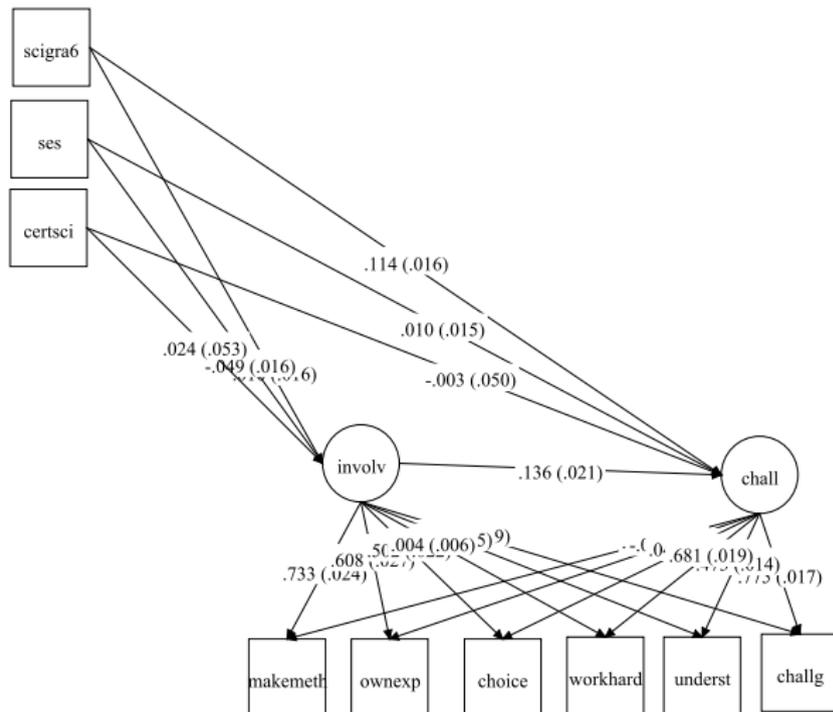
which is exactly the same as the Geomin rotation criterion.

- The first argument determines the loading matrix dimensions (the number of factors in the EFA). The second argument acts as penalty weight $w = 1/\nu$. The third argument is optional and is defaulted to 0.01.
- The order of the loading parameters is important

- Introduction to PSEM
- Univariate Priors
- PSEM - growth model
- Multivariate priors
- **PSEM-EFA example**
- Arbitrary prior/penalty
- MIMIC model with direct effects
- LASSO for the residual covariances in EFA
- Exploratory Latent Growth Models
- Alignment Illustration

- Example from Muthén & Asparouhov (2012)
- 2 EFA factors are measured by 6 indicators and we have 3 covariates.
- We want to regress factors 2 on factor 1 - this can only be done with PSEM-EFA
- Covariates predict factor 1 and we want to see if there are direct effects to factor 2
- The null model is the unrotated EFA but we can use the standard EFA with covariates to check log-likelihood

EFA example



EFA example input file

In the Geomin penalty 2 is the number of factors, 1 is 1/penalty weight, 0.0001 is the epsilon. Order in loading names is important.

```
model:
```

```
involv by makemeth*1 ownexp*1 choice*1 challg*0 underst*0 workhard*0 (p1-p6);  
chall by makemeth*0 ownexp*0 choice*0 challg*1 underst*1 workhard*1 (p7-p12);  
involv@1 chall@1;
```

```
chall on involv;  
involv on scigra6-certsci;  
chall on scigra6-certsci*0;
```

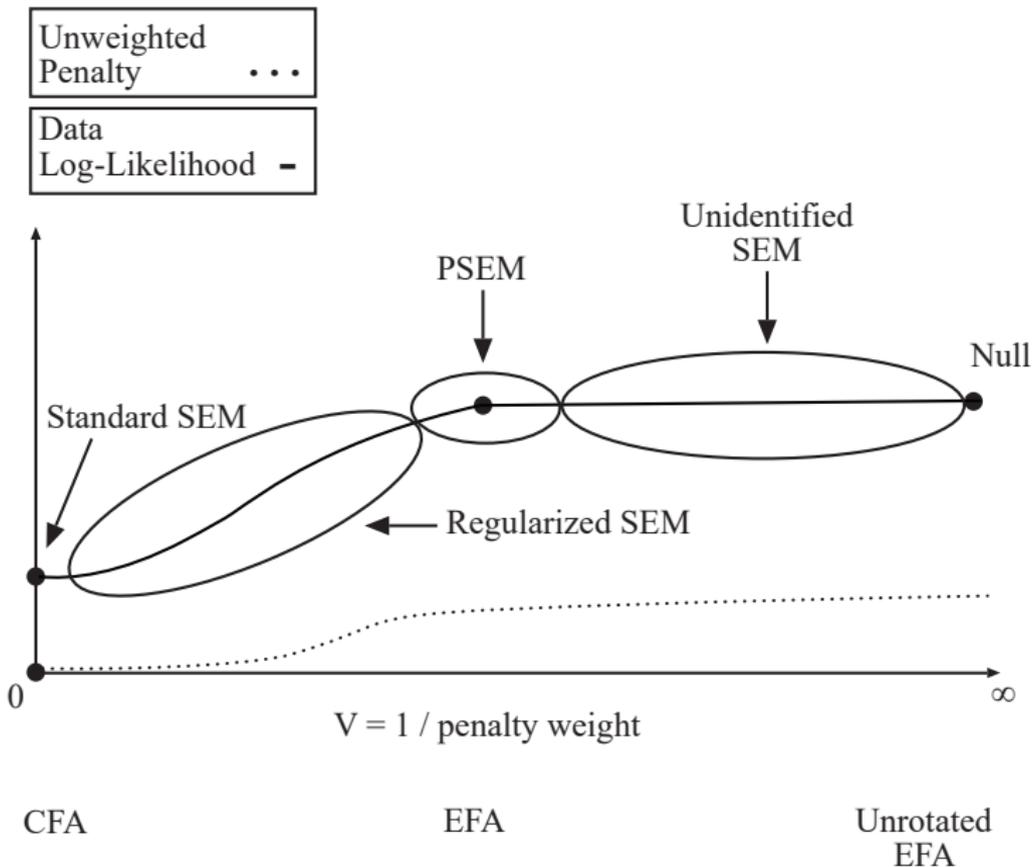
```
analysis: iter=10000; starts=50; conv=0.000001;
```

```
model priors: p1-p12~GEOMIN(2,1,0.0001);
```

EFA general observations

- There are **no** Geomin settings that produce exactly the same as a standard rotation in all situations
- The first argument is just the number of factors
- The second argument is where experimentation is needed but we just use powers of 10: 0.01, 0.1, 1, 10, 100
- You want to use the smallest of these values that gives approximately the same log-likelihood as the null model.
- Some structured ESEM-PSEM models don't have a null model and may not need Geomin prior and the rotation might be identified from the model and data: native rotation (ex different predictors for EFA factors may identify loading rotation and can give better fit)
- The third setting: in ESEM by default is 0.0001 for $m=2$, 0.001 for $m=3$, and 0.01 for bigger m . Geomin prior can be set that way but is defaulted at 0.01
- Many times you need larger ITER, smaller CONV, larger STARTS, and larger or smaller third Geomin argument

PSEM penalty weight v.s. Log-likelihood



- Introduction to PSEM
- Univariate Priors
- PSEM - growth model
- Multivariate priors
- PSEM-EFA example
- **Arbitrary prior/penalty**
- MIMIC model with direct effects
- LASSO for the residual covariances in EFA
- Exploratory Latent Growth Models
- Alignment Illustration

- Arbitrary penalty function can be constructed as follows. In `MODEL CONSTRAINT` define a new parameter `Penalty`. Then give the penalty a `LASSO(0,1)` prior and set the `TOLERANCE` option to 0 if `Penalty` is always positive.
- It helps to use `DO` and `SUM` operations in `MODEL CONSTRAINT` to make complex penalty functions
- Examples of penalty functions that might be of interest is rotation criteria different from `GEOMIN` such as `VARIMAX`, `QUARTIMAX` and some growth model rotations

PSEM: Arbitrary penalty function example

- This example shows how one can define manually the Quartimin rotation criterion. For orthogonal rotation add F1 with F2@0. For oblique rotation add (not necessary since it is by default) F1 with F2*0

```
ANALYSIS: TOLERANCE=0
```

```
MODEL:
```

```
  f1 BY y1*1 y2-y20 (a1-a20); f1@1;
```

```
  f2 BY y1*1 y2-y20 (b1-b20); f2@1;
```

```
MODEL CONSTRAINT:
```

```
  new(a); new(c1-c20);
```

```
  DO(#,1,20) c#=(a#*b#)^2;
```

```
  a=SUM(c1-c20);
```

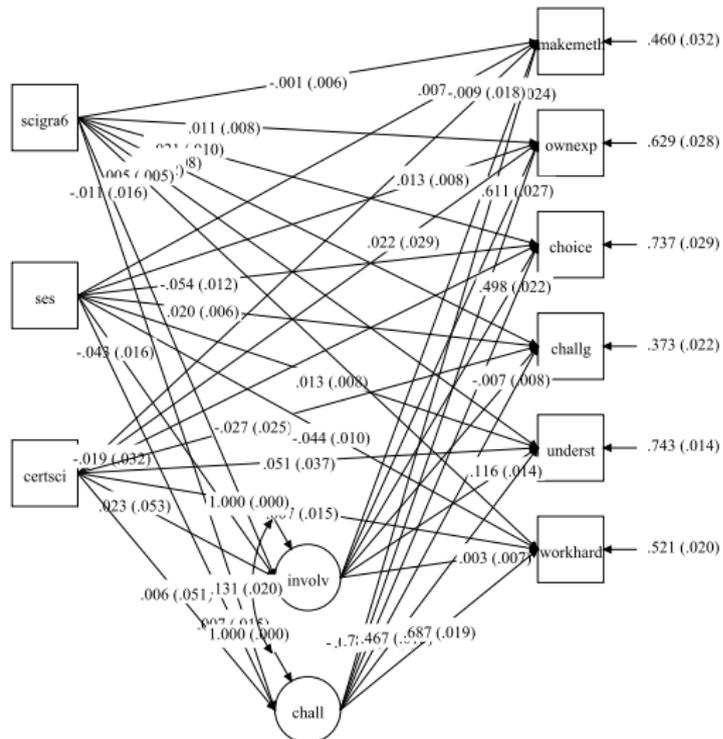
```
MODEL PRIORS: a~LASSO(0,1)
```

- Introduction to PSEM
- Univariate Priors
- PSEM - growth model
- Multivariate priors
- PSEM-EFA example
- Arbitrary prior/penalty
- **MIMIC model with direct effects**
- LASSO for the residual covariances in EFA
- Exploratory Latent Growth Models
- Alignment Illustration

- Example from Muthén & Asparouhov (2012) EFA with 2 factors, we want to see if there are direct effects to the indicators from the covariates that can not be explained by the effects on the factors
- Estimating all direct effects and all effects to the factors is an unidentified model
- A maximum of $6(\text{indicator}) * 3(\text{covariates})$ sufficient statistics are available and can not identify 24 parameters
- PSEM identifies the main effects to the factors, in addition to the direct effects, by imposing a penalty function which says that the direct effects should be small or zero as in LASSO regression

```
model:  
involv by makemeth-workhard (*1);  
chall by makemeth-workhard (*1);  
  
involv chall on scigra6-certsci;  
makemeth-workhard on scigra6-certsci* $\theta$  (d1-d18);  
  
analysis: iter=10000; starts=50; conv=0.000001;  
  
model priors: d1-d18~LASSO(0,1);
```

MIMIC model with direct effects diagram



- The model produces the same log-likelihood and number of parameters as the null model (no effects on factors)
- Two factor effects are found only and are pretty much the same as if you run the model without the direct effects
- LASSO and ALF give nearly identical results, LASSO finds 6 out of 18 significant direct effects, ALF finds 5 out of 18.
- LASSO direct effects by absolute value total is .395 vs. ALF .384 v.s. the null .620
- The PSEM model provides 36% reduction in direct effects without compromising the log-likelihood by channeling the effects through the factors

- Introduction to PSEM
- Univariate Priors
- PSEM - growth model
- Multivariate priors
- PSEM-EFA example
- Arbitrary prior/penalty
- MIMIC model with direct effects
- **LASSO for the residual covariances in EFA**
- Exploratory Latent Growth Models
- Alignment Illustration

LASSO for the residual covariances in EFA

- In an EFA model we can add all residual covariances with LASSO or ALF prior and still estimate the EFA model
- Such an approach would identify correlated uniqueness among the indicators that are not explained by the EFA

MODEL:

```
F1-F4 BY qcmt101-qcmt111 (*1);
```

```
qcmt101-qcmt111 with qcmt101-qcmt111 (p1-p55);
```

```
model prior: p1-p55~LASSO(0,1);
```

- PSEM is much easier to specify than Bayes IW prior which also involves variances
- The null model is the H1 model so DF will always be 0, but log-likelihood and fit function are available to monitor fit progress as LASSO weight is relaxed.
- As with BSEM the model provides a continuum between the H1 model and the EFA model
- Strategy for penalty weight: Start with a weight that yields exact fit as EFA, then increase variance gradually so residual misfit pops up. Use a follow-up model.
- 3 x 3 block means you need another factor.

- EFA with 4 factors yields chi-2 of 65, DF=17, p-val=.00
- EFA with LASSO $\nu=0.0001$ yields exact fit as EFA
- EFA with LASSO $\nu=0.01$ residual covariances pop-up
- Pick top 2 values (by Z-test values) (7,10) and (7,11)
- EFA with 2 residual covariances chi-2 23, DF=15, p-val=.07
- In complex examples p-value can be substituted with CFI/TLI

- Introduction to PSEM
- Univariate Priors
- PSEM - growth model
- Multivariate priors
- PSEM-EFA example
- Arbitrary prior/penalty
- MIMIC model with direct effects
- LASSO for the residual covariances in EFA
- **Exploratory Latent Growth Models**
- Alignment Illustration

Exploratory Latent Growth Model(PSEM-ELGM)

- ELGM started by Tucker (1958, 1966) and more recently in Grimm et al. (2013) with ESEM
- In standard growth models the loadings are fixed. In exploratory growth model all loadings are free

$$Y = \Lambda\eta + E$$

$$\eta \sim N(\alpha, \Sigma)$$

- This model is different from the standard growth model because the loading matrix is estimated. In standard growth models the loading matrix is fixed

$$\Lambda^T = \begin{pmatrix} 1 & 1 & 1 & \dots \\ 0 & 1 & 2 & \dots \end{pmatrix}$$

- This model is different from EFA because Y doesn't have an intercept parameter while η has
- The EFA loadings structure is entangled with the mean structure

- Grimm et al. (2013) finds that "new specialized rotation criteria" are needed.
- EFA aims for simple loading structures, every indicator loads on one factor, which does not resemble growth curve and is difficult to interpret
- Growth models rotation criteria needs the opposite: random intercept that loads on all indicators, i.e., baseline status
- Developmental/growth factors progressively loads on later indicators
- PSEM has the tools needed to build custom growth model rotations

- If we want to model the first factor as a random intercept, we model λ_{t1} as a constant, $DIFF(\lambda_{11} - \lambda_{T1}) \sim ALF(0, 1)$
- To model linear loadings, λ_{t2} would be a $c \cdot t$, with a constant first derivative, i.e., $\lambda'_{t2} = \lambda_{t+1,2} - \lambda_{t,2}$,
 $DIFF(\lambda'_{12} - \lambda'_{T-1,2}) \sim ALF(0, 1)$
- SEM based variations of this model are commonly used: instead of fixing the loadings at time scores some loadings are free parameters. ELGM generalizes this method.

- In addition to exploratory curve modeling we can also add time specific means as was done earlier

$$Y_{it} = \nu_t + \lambda_{1t}I_i + \lambda_{2t}S_i + \varepsilon_{it} \quad (5)$$

$$Y_{it} = \alpha(t) + \beta(t)I_i + \gamma(t)S_i + \varepsilon_{it} \quad (6)$$

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

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

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

- The estimated curve is a mixture of three curves: population based time adjustment $\alpha(t)$, baseline curve $\beta(t)$ and growth curve $\gamma(t)$
- I_i and S_i are subject specific weights for $\beta(t)$ and $\gamma(t)$

ELGM uses BY statement and not | like standard growth models.

```
model:
  i by pa1-pa7*1 (a1-a7);
  s by pa1*0 pa2*.1 pa3*.2 pa4*.3 pa5*.4 pa6*.5 pa7*.6 (b1-b7);
  [pa1-pa7] (m1-m7);
  pa2^-pa7^ pon pa1^-pa6^ (ar);
  i@1 s@1; [i s];

model prior:
  m1-m7~ALF(0,1);
  DIFF(a1-a7)~ALF(0,10);
  b1~ALF(0,1);
```

ELGM Rotation criterion

- For ELGM with intercept and slope the above criteria appears to work well universally
- $m1 - m7 \sim ALF(0, 1)$; identifies $\alpha(t)$ and the factor means
- $DIFF(a1 - a7) \sim ALF(0, 10)$; forces the I loadings to be approximately the same so that we can interpret the random effect as a random intercept. Using a second argument of 10 instead of 1 is to lessen the strength of the penalty. Such an adjustment is fairly common and is explained by the fact that the DIFF prior produces 21 terms, not just 7, because it includes all pairwise comparisons
- ELGM has been used with consecutive differences in the literature instead of pairwise differences. We prefer pairwise as it corresponds to the Bayes multivariate normal DIFF prior where a group of parameters have prior correlation of .99
- $b1 \sim ALF(0, 1)$; is simply choosing the timescale as to where 0 is and corresponds to fixing the loading to 0 at the first time

- The null model is the EFA model (with auto-regression), i.e., we must aim at getting that EFA fit
- The null with auto-regression gives Heywood case. Without auto-correlation is ok so we proceed without it (or hold negative residual equal to neighbouring time point residual)
- PSEM-ELGM log-likelihood -1303.563, EFA log-likelihood -1303.452. Sufficiently close. No need to modify the penalty weight further.

PSEM-ELGM Results

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
I	BY				
	PA1	0.723	0.048	15.082	0.000
	PA2	0.906	0.049	18.676	0.000
	PA3	0.899	0.049	18.329	0.000
	PA4	0.911	0.049	18.440	0.000
	PA5	0.739	0.062	11.994	0.000
	PA6	0.901	0.047	19.017	0.000
	PA7	0.895	0.048	18.687	0.000
S	BY				
	PA1	-0.000	0.000	-0.818	0.413
	PA2	-0.052	0.080	-0.658	0.511
	PA3	0.061	0.083	0.738	0.461
	PA4	0.211	0.091	2.310	0.021
	PA5	0.423	0.125	3.377	0.001
	PA6	0.726	0.113	6.411	0.000
	PA7	0.516	0.095	5.422	0.000
Intercepts					
	PA1	1.172	0.314	3.730	0.000
	PA2	-0.009	0.037	-0.234	0.815
	PA3	-0.000	0.037	-0.001	0.999
	PA4	0.011	0.048	0.233	0.816
	PA5	1.231	0.306	4.027	0.000
	PA6	0.003	0.025	0.136	0.892
	PA7	-0.017	0.054	-0.309	0.757

ELGM Results

- I* can be interpreted as an approximate random intercept.
Detailed invariance analysis is available with OUTPUT:ALIGN

DIFF ANALYSIS FOR PARAMETERS

A1	A2	A3	A4	A5	A6	A7
Chi-square value			34.112			
Degrees of freedom			6			
P-value			0.000			

Approximate Invariance Holds For:

A2	A3	A4	A5	A6	A7
Average Value Across Invariant Parametetrs:					0.875

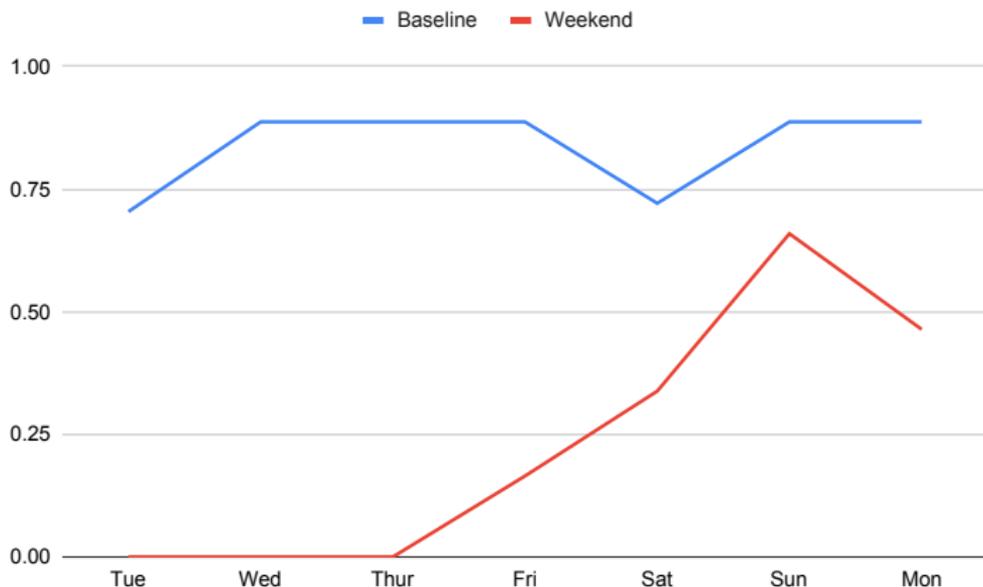
Invariant Values, Difference to Average and Significance

Param	Value	Difference	SE	P-value
A2	0.906	0.031	0.046	0.498
A3	0.899	0.024	0.046	0.607
A4	0.911	0.036	0.046	0.437
A5	0.739	-0.137	0.060	0.022
A6	0.901	0.026	0.044	0.561
A7	0.895	0.020	0.045	0.655

- S can be interpreted as a weekend effect which culminates on Sunday
- The final PSEM-ELGM follow up model

```
model:
  i by pa1*1 pa5*1;
  i by pa2*1 pa3*1 pa4*1 pa6*1 pa7*1 (1);
  s by pa1-pa3@0 pa4*.3 pa5*.4 pa6*.5 pa7*.6;
  [pa1 pa5 pa2-pa4@0 pa6-pa7@0 ];
  pa2^-pa7^ pon pa1^-pa6^ (ar);
  i@1 s@1; [i s@0];
```

PSEM-ELGM Results



Individual curves are based on adding up these two curves with individually specific weights (the factors are the weights), plus two universal parameters for Tue and Sat). Weight range: Blue(4.5,8.5) Red(-2,2).

Table: PSEM PA growth model

Model	LL	Num Param	Chi-2	p-val	BIC
Cubic	-1318	15	31	.08	2720
PSEM LG	-1313	18	21	.23	2725
PSEM LG Null	-1313	18	21	.23	2725
PSEM LG follow-up	-1318	13	29	.14	2707
PSEM RI	-1321	16	34	.02	2729
PSEM RI follow-up	-1325	11	43	.01	2710
ELGM / EFA	-1304	27	7	.53	2756
ELGM follow-up	-1305	19	13	.68	2715

- At $N = 244$ perhaps ELGM style modeling is not justified. Bigger samples are needed to gain numerical advantage.
- ELGM can offer superior and more realistic interpretation. Strictly linear increase in PA is not realistic
- ELGM can be used to provide a more refined growth model in the space between EFA and the standard growth models
- ELGM can improve model fit
- ELGM can improve random effect predictions

- Introduction to PSEM
- Univariate Priors
- PSEM - growth model
- Multivariate priors
- PSEM-EFA example
- Arbitrary prior/penalty
- MIMIC model with direct effects
- LASSO for the residual covariances in EFA
- Exploratory Latent Growth Models
- **Alignment Illustration**

PSEM alignment application

- Longitudinal alignment for one factor measured at T=8 time points by 3 ordered categorical variables
- N=1174. Items are measuring disruptive behavior in the classroom, 1-th to 7-th grade (two observations in 1-th grade): Break Things, Harm Others and Property, Take Others' Property
- The three variables have 7 categories each and are highly skewed: category 1 has approximately 70% of the observation, second category has approximately 10-20%, third around 5% and single digits or less for higher categories.
- The highest categories are so rare that not all occur at all time points.
- The automatic alignment procedure can not be used because the number of thresholds varies across time (although one can eliminate rare categories by combining with other categories)
- Even if all categories occur at all time points, manual PSEM-alignment procedure is more illuminating (rather than a black box automatic procedure)

- Rare high categories result in volatile high threshold values that may unduly influence the alignment
- We will use the WLSMV estimator which provides a chi-square test of fit with categorical variables
- PSEM currently unavailable with ML estimator for categorical variables, and ML can not estimate 8 dimensional integration anyway
- Outline for Alignment:
 - PSEM metric alignment
 - Additions to metric alignment to obtain a good correlation fit
 - PSEM scalar alignment

- We start by evaluating metric invariance via PSEM first, alignment of loadings and estimating factor variances.
- The first step is to evaluate the null model: the configural invariance model
- There are two parameterizations in Mplus: theta and delta.
- The configural model is equivalent in both, but the metric and scalar invariance models are not (with or without alignment)

```
model:  
f1 by bkthin1f*1 harmo1f takeP1f;  
f2 by bkthin1s*1 harmo1s takeP1s;  
f3 by bkthin2s*1 harmo2s takeP2s;  
f4 by bkthin3s*1 harmo3s takeP3s;  
f5 by bkthin4s*1 harmo4s takeP4s;  
f6 by bkthin5s*1 harmo5s takeP5s;  
f7 by bkthin6s*1 harmo6s takeP6s;  
f8 by bkthin7s*1 harmo7s takeP7s;  
f1-f8@1;
```

Configural model results

MODEL FIT INFORMATION

Number of Free Parameters 175

Chi-Square Test of Model Fit

Value	356.847*
Degrees of Freedom	224
P-Value	0.0000

- * The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used for chi-square difference testing in the regular way. MLM, MLR and WLSM chi-square difference testing is described on the Mplus website. MLMV, WLSMV, and ULSMV difference testing is done using the DIFFTEST option.

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.022	
90 Percent C.I.	0.017	0.026
Probability RMSEA <= .05	1.000	

CFI/TLI

CFI	0.994
TLI	0.993

Configural model results

Chi-Square Test of Model Fit for the Baseline Model

Value	22998.342
Degrees of Freedom	276
P-Value	0.0000

SRMR (Standardized Root Mean Square Residual)

Value	0.031
-------	-------

Optimum Function Value for Weighted Least-Squares Estimator

Value	0.80562082D-01
-------	----------------

- The chi-square rejects the model but the configural model provides an excellent approximate fit according to CFI/TLI
- With Alignment we can't get a better fit than the configural model, the fit should be the same
- The fit function for WLSMV does not have a chi-square distribution and is rescaled to obtain the chi-square statistic
- Direct testing between two nested models can not be done using only the chi-square value. In Mplus a special utility can be used called DIFFTEST for testing nested models
- The chi-square alone can not be used for monitoring the distance between a penalized model and a null model
- We must use **Optimum Function Value for Weighted Least-Squares Estimator** instead of chi-square
- The penalty should be low enough as to get at least 3 or 4 top significant digits

PSEM metric alignment delta model

```
model:
f1 by bkthin1f*1 harmo1f takeP1f (a1_1-a1_3);
f2 by bkthin1s*1 harmo1s takeP1s (a2_1-a2_3);
f3 by bkthin2s*1 harmo2s takeP2s (a3_1-a3_3);
f4 by bkthin3s*1 harmo3s takeP3s (a4_1-a4_3);
f5 by bkthin4s*1 harmo4s takeP4s (a5_1-a5_3);
f6 by bkthin5s*1 harmo5s takeP5s (a6_1-a6_3);
f7 by bkthin6s*1 harmo6s takeP6s (a7_1-a7_3);
f8 by bkthin7s*1 harmo7s takeP7s (a8_1-a8_3);
f1@1; f2-f8*1;

model prior:
DO(#,1,3) DIFF(a1_# a2_# a3_# a4_# a5_# a6_# a7_# a8_#)~ALF(0,1);

output: align;
```

- Using ALF variance/weight of 1 we get a fit function value of 0.080680402 v.s. null model 0.080560830
- Using ALF variance/weight of 10 (lower weight) we get a fit function value of 0.80562082 very close (sufficiently close)
- The chi-square for this model is 356.847 v.s. 357.237 for the null model. Close but note the reverse order and emphasize again fit function use instead of chi-square
- Next we evaluate the new parameters: factor variances.

PSEM metric alignment delta model

Variances

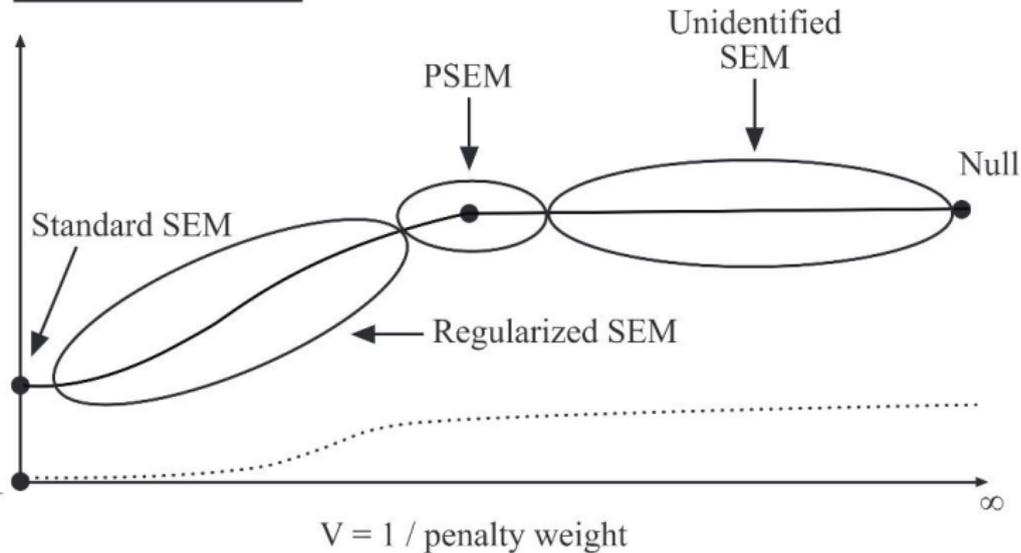
F1	1.000	0.000	999.000	999.000
F2	0.898	0.023	39.528	0.000
F3	0.940	0.023	41.331	0.000
F4	0.959	0.027	35.230	0.000
F5	0.919	0.034	27.152	0.000
F6	0.931	0.026	36.310	0.000
F7	0.959	0.027	35.704	0.000
F8	0.907	0.028	32.011	0.000

- Factor variances are well identified and are statistically different from 1 which means that are needed
- If the penalty weight is too low, the SE would be too large as compared to the metric invariance model
- We pick the largest penalty weight (lowest ALF variance) that gave the same fit function as the null model
- The prior proportion is 0.02 which is acceptable (safe range 0.01 to 0.2)
- The penalty weight provides a continuum between the configural model and the metric model

Continuum

Unweighted
Penalty ...

- Fit function_



Metric invariance

Metric Alignment

Configural

PSEM metric alignment delta model results

Next we evaluate loading invariance in OUTPUT:ALIGN

DIFF ANALYSIS FOR PARAMETERS

A1_1	A2_1	A3_1	A4_1	A5_1	A6_1	A7_1	A8_1
------	------	------	------	------	------	------	------

Chi-square value			11.829				
------------------	--	--	--------	--	--	--	--

Degrees of freedom			7				
--------------------	--	--	---	--	--	--	--

P-value			0.106				
---------	--	--	-------	--	--	--	--

DIFF ANALYSIS FOR PARAMETERS

A1_2	A2_2	A3_2	A4_2	A5_2	A6_2	A7_2	A8_2
------	------	------	------	------	------	------	------

Chi-square value			5.812				
------------------	--	--	-------	--	--	--	--

Degrees of freedom			7				
--------------------	--	--	---	--	--	--	--

P-value			0.562				
---------	--	--	-------	--	--	--	--

DIFF ANALYSIS FOR PARAMETERS

A1_3	A2_3	A3_3	A4_3	A5_3	A6_3	A7_3	A8_3
------	------	------	------	------	------	------	------

Chi-square value			4.388				
------------------	--	--	-------	--	--	--	--

Degrees of freedom			7				
--------------------	--	--	---	--	--	--	--

P-value			0.734				
---------	--	--	-------	--	--	--	--

PSEM metric invariance model results

- All 3 Wald tests imply metric invariance holds in the default delta parametrization
- Next we repeat the estimation in the theta parameterization with ANALYSIS: PARAM=THETA;
- Using ALF variance of 1 and 10 we did not match the null fit function. Using 100 we did.
- To get convergence with 100 we needed to increase ITER=10000. Not unusual. The error message says "Number of iterations exceeded".
- In the THETA parameterization Wald test for the first indicator has a p-value of 0.024
- Further detailed analysis (below Wald test) shows that at T=7, the first loading is 1.3 and the average of all other time points is 2.0
- Conclusion: metric invariance holds with the delta parameterization but not theta and we would want to continue the analysis with the delta parameterization

- Chi-square test of fit for the two metric models is not very different 370 vs 372 (with the same number of parameters 161) and thus we can keep track of the fit for both.
- The two metrics have different fit when the factor variances are statistically significantly different from 1. At this moderate sample size, not a big difference between the models (typical)
- Before we proceed with scalar invariance, we hold the mean structure unrestricted and we attempt to improve the correlations fit (separation of mean and correlation structure).
- Improvement 1: Time invariant but time specific correlated uniqueness among the three indicators via PSEM

PSEM metric invariance model

```
Analysis: param=theta; conv=0.000001; iter=10000;
```

```
model:
```

```
f1 by bkthin1f*1 harmo1f takeP1f (a1-a3);  
f2 by bkthin1s*1 harmo1s takeP1s (a1-a3);  
f3 by bkthin2s*1 harmo2s takeP2s (a1-a3);  
f4 by bkthin3s*1 harmo3s takeP3s (a1-a3);  
f5 by bkthin4s*1 harmo4s takeP4s (a1-a3);  
f6 by bkthin5s*1 harmo5s takeP5s (a1-a3);  
f7 by bkthin6s*1 harmo6s takeP6s (a1-a3);  
f8 by bkthin7s*1 harmo7s takeP7s (a1-a3);  
f1@1; f2-f8*1;
```

```
bkThin1f-bkThin7s pwith takeP1f-takeP7s (v1);  
harm01f-harm07s pwith takeP1f-takeP7s (v3);  
harm01f-harm07s pwith bkThin1f-bkThin7s (v2);
```

```
model prior: v1-v3~ALF(0,1);
```

- We use the theta parameterization where the residual variances are fixed to 1 and the new parameters represent correlation
- With the delta parameterization it would not be a time invariant correlation but a time invariant covariance with non-invariant variances so not as natural
- Without the prior the model is unidentified
- With the prior the model is identified (proper PSEM).
- With the prior the model has 163 parameters v.s. 161 for the metric model, i.e., 2 of the three parameters are identifiable but not all 3 of them
- Establishing this fact without Mplus would require complicated matrix algebra.
- This is also not an intuitive result. Without the longitudinal settings neither one of the three parameters is identified

- Using ALF with variance 1 yields chi-square of 314 with NP=163 vs the metric invariance model of 372 with NP=161
- Using ALF with variance 10 yields the same chi-square and fit function so we can rely on ALF(0,1) results (prior ratio is 0.01)
- Two of the three parameters are significant.
- PSEM follow-up model: Remove PSEM (prior/penalty) but keep the two parameters. This yields a chi-square of 314 and NP=163.

- Next we attempt three additional modifications: Random intercepts (independent/residual) for each of the three indicators, methods factors (same as RI but with free loadings), time invariant residual auto-regression for the three indicators
- Only methods factor makes a bigger impact: chi-square 252, NP=187
- Next we structure the factor variance covariance: random intercept with free loadings(second order factor) with auto-regressive residual: chi-square 258 NP=174, p-val=.06. This will be the null model for the scalar invariance
- Next we proceed with evaluating scalar invariance, alignment of thresholds and estimating the factor means

PSEM scalar invariance alignment

```
model:
! metric invariance
f1 by bkthin1f*1 harmo1f takeP1f (a1-a3);
f2 by bkthin1s*1 harmo1s takeP1s (a1-a3);
f3 by bkthin2s*1 harmo2s takeP2s (a1-a3);
f4 by bkthin3s*1 harmo3s takeP3s (a1-a3);
f5 by bkthin4s*1 harmo4s takeP4s (a1-a3);
f6 by bkthin5s*1 harmo5s takeP5s (a1-a3);
f7 by bkthin6s*1 harmo6s takeP6s (a1-a3);
f8 by bkthin7s*1 harmo7s takeP7s (a1-a3);
f1@1; f2-f8*1; [f2-f8*0]; !<-free factor means

! random intercept with AR for the factors
f by f1-f8*1; f@1; f2^-f8^ pon f1^-f7^;

! time specific correlated uniqueness
bkThin1f-bkThin7s pwith takeP1f-takeP7s (v1);
harmO1f-harmO7s pwith bkThin1f-bkThin7s (v2);

! random intercepts for each indicator
i1 by bkThin1f-bkThin7s*1; i1@1;
i2 by harmO1f-harmO7s*1; i2@1;
i3 by takeP1f-takeP7s*1; i3@1;
i1-i3 with i1-i3@0; f1-f8 with i1-i3@0; f with i1-i3@0;
```

PSEM scalar invariance alignment

```
! giving names to the threshods
[bkthin1f$1 harmo1f$1 takeP1f$1] (a1_1-a1_3);
[bkthin1s$1 harmo1s$1 takeP1s$1] (a2_1-a2_3);
[bkthin2s$1 harmo2s$1 takeP2s$1] (a3_1-a3_3);
[bkthin3s$1 harmo3s$1 takeP3s$1] (a4_1-a4_3);
[bkthin4s$1 harmo4s$1 takeP4s$1] (a5_1-a5_3);
[bkthin5s$1 harmo5s$1 takeP5s$1] (a6_1-a6_3);
[bkthin6s$1 harmo6s$1 takeP6s$1] (a7_1-a7_3);
[bkthin7s$1 harmo7s$1 takeP7s$1] (a8_1-a8_3);

[bkthin1f$2 harmo1f$2 takeP1f$2] (b1_1-b1_3);
[bkthin1s$2 harmo1s$2 takeP1s$2] (b2_1-b2_3);
[bkthin2s$2 harmo2s$2 takeP2s$2] (b3_1-b3_3);
....
```

model prior:

```
DO(#,1,3) DIFF(a1_# a2_# a3_# a4_# a5_# a6_# a7_# a8_#)~ALF(0,1);
DO(#,1,3) DIFF(b1_# b2_# b3_# b4_# b5_# b6_# b7_# b8_#)~ALF(0,1);
DO(#,1,3) DIFF(c1_# c2_# c3_# c4_# c5_# c6_# c7_# c8_#)~ALF(0,1);
DO(#,1,3) DIFF(d1_# d2_# d3_# d4_# d5_# d6_# d7_# d8_#)~ALF(0,1);
DO(#,1,3) DIFF(e1_# e2_# e3_# e4_# e5_# e6_# e7_# e8_#)~ALF(0,1);
```

PSEM scalar invariance alignment

Model	Fit Func	Chi-2	P-val	NP	Wald	Penalty
Metric (Null)	.0695	258	.061	174	328	257
V=100 (PSEM)	.0695	258	.061	174	166	199
V=10	.0701	257	.066	174	178	155
V=1	.1022	286	.003	174	129	56
V=.1	.1471	364	.000	174	370	14
Scalar	.1484	492	.000	76	0	0

- Wald test is not a very good measure of evaluating overall alignment in the thresholds as the PSEM models with the extra parameters have larger SE
- Penalty Diff is unweighted and more accurately reflects the threshold alignment explained by factor means - here only 23% reduction in threshold difference was archived with PSEM (V=100)
- V provides a continuum between Metric and Scalar and PSEM is on that continuum.

PSEM scalar invariance alignment

- Wald Diffs are computed in OUTPUT:ALIGN and are manually added for the 3(variables)*5(thresholds)
- The Penalty Diff are computed using the following formula

$$PenaltyDiff = [PriorProportion] * [FitFunc] * N * V$$

- For the scalar model these are automatically 0
- For the Null model one must estimate the null model as a PSEM model and apply the above formula.
- To do that we simply use $V = 1000000$ (a very large value). With such a huge V the penalty weight is basically zero and at that point the factor means will not be identified.
- In this PSEM estimation the factor means must be fixed to zero as they are in the Null model.
- Note that here [Prior Proportion] will be a very small number and V will be very large but these will cancel each other out.
- This is just a trick to engage the PSEM code in Mplus to produce the Penalty Diff or the Wald tests for the Null model

PSEM metric invariance model

- The aligned PSEM model ($V=100$) is identified on the continuum (V from 0 to infinity) between scalar and metric as the first model which gets the same fit as metric and by definition it achieves the smallest penalty diff conditional on metric data fit
- The model can also be converted to the delta parameterization with similar success
- This alignment does not allow Free alignment which only works when there is metric non-invariance of sufficient size
- One question is if the PSEM model with $V=10$ is of value. The answer is probably yes. Fit function has a minor difference only and is not rejected. The model may offer smaller standard errors for some key parameters. Interpretation of the model is the same. Model prior is another way to write model constraints. If we want to impose smaller threshold variation (40% reduction in Penalty Diffs v.s. 23%) and accept a slightly worse model fit.
- This model won't be a PSEM model but regularized SEM, since we sacrificed some data fit for better alignment

- We want to add gender as a predictor to the above model.
- PSEM allows us to estimate: effect on the main factor, effect on each grade specific factors and on each measurement at every time point
- The null model is the model where all indicators are regressed on gender. That model does not allow us to evaluate the gender effect on the factors because it becomes unidentified

```
model:  
...  
f on gender;  
f1-f8 on gender (b1-b8);  
bkThin1f - takeP7s on gender (d1-d24);  
  
model prior:  
....  
b1-b8~ALF(0,10);  
d1-d24~ALF(0,10);
```

PSEM MIMIC structural alignment

- The null model chi-square is 257, NP=198, pval=0.066, Fit func=0.072092
- The PSEM model chi-square is 258, NP=198, pval=0.063, Fit func=0.072109
- If we only regress the main factor on gender (standard MIMIC), the chi-square is 289, NP=175, pval=0.037, Fit func=0.085876
- PSEM model produces nearly identical main factor effect, but also produces gender effect specific to each grade and effect specific to each item
- The goal of PSEM is not to pursue chi-square p-values but to extract maximum amount of information that is available in the data
- The model is called structural alignment because it aligns the gender effects along the MIMIC model where the main effect is on the main factor and all other effects are treated as residual effects

PSEM MIMIC structural alignment results

F ON GENDER	-1.014	0.238	-4.266	0.000
F1 ON GENDER	0.043	0.195	0.219	0.827
F2 ON GENDER	-0.117	0.193	-0.610	0.542
F3 ON GENDER	0.084	0.146	0.574	0.566
F4 ON GENDER	0.007	0.035	0.210	0.833
F5 ON GENDER	0.154	0.191	0.806	0.420
F6 ON GENDER	-0.050	0.189	-0.264	0.792
F7 ON GENDER	-0.099	0.188	-0.525	0.599
F8 ON GENDER	-0.103	0.178	-0.580	0.562
BKTHIN1F ON GENDER	0.015	0.071	0.207	0.836
BKTHIN1S ON GENDER	0.000	0.036	-0.002	0.998
BKTHIN2S ON GENDER	-0.012	0.076	-0.159	0.873

PSEM alignment conclusions

- There is value in keeping the PSEM-alignment model as the last model
- It is difficult in practice to move from an alignment model to a follow up model where the penalty is removed and just non-invariant parameters are kept as free
- OUTPUT: align is not always easy to utilize to build a follow up model
- Keep the alignment PSEM model to get the benefit of being able to estimate factor means and variances without enforcing scalar invariance when it doesn't hold and when it is messy to identify non-invariant parameters.
- The same applies to structural alignment.
- Simulation studies provide an unrealistic picture where the level of non-invariance is too low as compared to what is in real data.

- More examples in Asparouhov, T., Muthén, B. (2023). Penalized structural equation models. Technical report.
<https://www.statmodel.com/download/PML.pdf>
- PSEM talk tomorrow
- PSEM is a masterful technique to see further in the data with bright future and potential for new methodological breakthroughs

Continuous time RDSEM

Continuous time RDSEM

- Suppose that Y_{ij} is a vector of observations for individual i at time t_{ij} , for $j = 1, \dots, N_i$, i.e., N_i observations are taken for individual i at times $t_{i1}, t_{i2}, \dots, t_{iN_i}$. We want to model this data with a subject specific VAR model
- Currently Mplus uses time discretization where the times of observations are approximated with a integer-like discrete grid using the TINTERVAL option
- The algorithm results in missing data being inserted in the data file for those occasions where no observation was taken
- For example, if observations are taken at hours 2 and 6, missing data is inserted at times 1,3,4,5, given hourly discretization
- If discretization is based on two-hour periods, then missing data is inserted in hour 4 only
- There are two problems: the missing data is imputed internally and if a large amount of missing values are inserted it can cause slow or no convergence

- The second problem is the absence of a native discretization interval. In daily diary journal data there is a natural time scale. Auto-correlation and cross-lag are defined by the effect of the previous day on the current day
- In the earlier example with hourly data, it is not clear which auto-correlation we should be looking at. And the estimates are different since the smaller the time interval the bigger the auto-correlation
- The size of the discrete grid time interval is specified with `TINTERVAL`
- Halving `TINTERVAL` acts as squaring the auto-regressive matrix

Continuous time RDSEM

- The continuous time RDSEM model provides a solution for both of these problem

$$Y_{ij} = YB_i + YW_{ij}$$

$$YW_{ij} = \beta_i XW_{ij} + Y\hat{W}_{ij}$$

$$Y\hat{W}_{ij} = R_i(\Delta t_{ij}) Y\hat{W}_{i,j-1} + \varepsilon_{ij}$$

$$\Delta t_{ij} = t_{ij} - t_{i,j-1}$$

$$R_i(\Delta t_{ij}) = \text{Exp}(\Delta t_{ij} \cdot \text{Drift}_i) \quad \text{matrix exponentiation}$$

- $R_i(\Delta t)$ represents the auto-regressive matrix between two consecutive observations: depends on the individual and the time distance between the two observations

- $Drift_i$ is subject specific but independent of the times of observations
- The vector of random effects $Drift_i$, β_i , and YB_i are between level normally distributed random effects modeled with a between level SEM which can include between level predictors
- A key concept is that unlike other Mplus models the within level variance parameter is the total variance Σ of $Y\hat{W}_{ij}$ and not the residual of the residual variance of ε_{ij} . Σ is independent of the time interval but the variance covariance of ε_{ij} depends on Δt :
 $\Sigma - R_i(\Delta t_{ij})\Sigma R_i(\Delta t_{ij})^T$
- The variance covariance of $Y\hat{W}_{ij}$ can be made individually specific as in regular DSEM/RDSEM via random effects for the transformed variance and correlation parameters
- If times of observations are regularly spaced the model reduces to the regular RDSEM model

- Caveat 1: Slow convergence, slower than RDSEM
- Caveat 2: Requires larger amount of data. For models with subject specific drift and residual variance covariance large amount of data needed for each individual: 50 and up
- Caveat 3: CT-RDSEM produces results that are very similar to TINTERVAL
- Caveat 4: With current development, only VAR model, not as general as RDSEM, lag 1 only, no missing data within Y_{ij}
- The model, however, is very attractive and complements DSEM in visualizing cross-lagged effects dependence on time. Possible novel applications
- Stay tuned - still searching for the best algorithm. Some of these caveats might be resolved

Continuous time RDSEM simulation example

```
MONTECARLO:  NAMES ARE y1-y2;
              NOBS = 1500;
              NREP = 100;
              NCSIZES = 1;
              CSIZES = 50(30);
              lagged=y1-y2(1);
              CTIME=IG(3.1,5.2);

ANALYSIS:    TYPE IS TWOLEVEL;
              estimator=bayes; proc=2; burnin=100;

MODEL MONTECARLO:

%WITHIN%
y1^ on y1^1*-0.5;
y2^ on y2^1*-0.6;
y1^ on y2^1*-0.2;
y2^ on y1^1*0.2;
y1*1.1; y2*1.3;
y1 with y2*0.5;

%BETWEEN%
y1*1.2 y2*0.5;
Y1 WITH Y2*0.4;
[Y1*3 Y2*0.3];

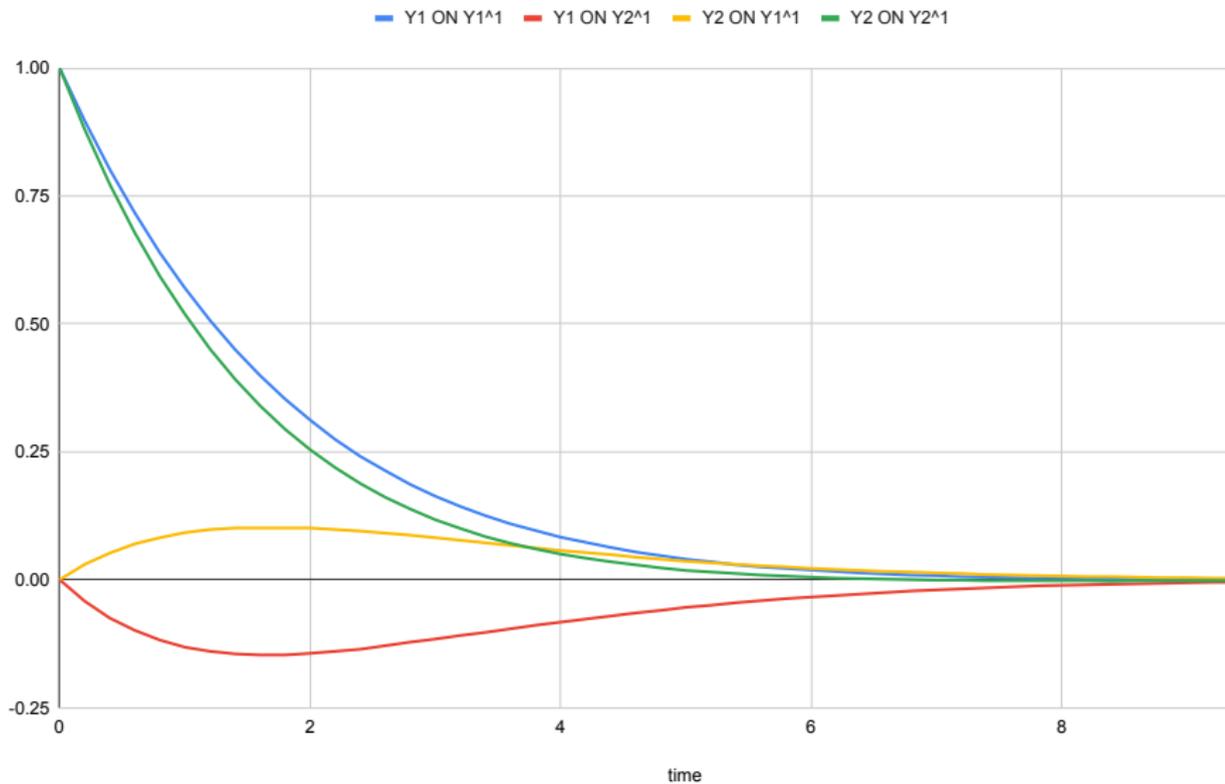
MODEL: (the same as generating model)
```

Continuous time RDSEM simulation example results

MODEL RESULTS

		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Within Level								
Y1	WITH							
Y2		0.500	0.4812	0.0381	0.0373	0.0018	0.910	1.000
Residual Variances								
Y1		1.100	1.0835	0.0451	0.0437	0.0023	0.880	1.000
Y2		1.300	1.2798	0.0622	0.0537	0.0042	0.870	1.000
Drift								
Y1^	ON							
Y1^1		-0.500	-0.5215	0.0456	0.0468	0.0025	0.960	1.000
Y2^1		-0.200	-0.2068	0.0482	0.0424	0.0024	0.910	0.990
Y2^	ON							
Y2^1		-0.600	-0.6114	0.0478	0.0416	0.0024	0.910	1.000
Y1^1		0.200	0.1891	0.0525	0.0471	0.0029	0.940	0.990
Between Level								
Y1	WITH							
Y2		0.400	0.4503	0.1525	0.1697	0.0255	0.930	0.970
Means								
Y1		3.000	3.0195	0.1578	0.1698	0.0250	0.980	1.000
Y2		0.300	0.3118	0.1094	0.1132	0.0120	0.960	0.750
Variances								
Y1		1.200	1.3571	0.2778	0.3217	0.1011	0.930	1.000
Y2		0.500	0.5592	0.1333	0.1444	0.0211	0.920	1.000

Visualizing drift exponentiation to auto-regressive effects



- The cross-lagged effects have a peaks: they always will
- The peak can be used to also determine optimal TINTERVAL value for DSEM and RDSEM to best illustrate cross-lagged effects
- Provides a clear illustration of how cross-lagged effects accumulate and dissipate over time