Dynamic Structural Equation Modeling of Intensive Longitudinal Data Using Mplus Version 8

Tihomir Asparouhov

Part 5

August 2, 2017

Overview

- The general DSEM model and estimation
- Model fit: DIC, model estimated means and variances
- Centering
- The MEAR model
- Covariates in DSEM, new RDSEM model (residual DSEM)

Reference for this talk is the ”Dynamic Structural Equation Models” paper available online http://statmodel.com/download/DSEM.pdf

The general DSEM model

- Merge ”time series”, ”structural equation”, ”multilevel” and ”TVEM(time varying effect modeling)” concepts in a generalized modeling framework in Mplus V8
- Y_{it}, η_{it} and X_{it} - are the observed dependent variables, latent factors and predictors for individual i at time t
- Four distinct sources of correlation in such observed data:
  - correlation due to individual specific effects (multilevel)
  - correlation due to proximity of observations (time series)
  - correlation between different variables (SEM)
  - correlation due to the same stage of evolution (TVEM)
- DSEM finds these correlations

DSEM model

- Includes three separate models: single level, twolevel , cross-classified
- Main decomposition equation

$$Y_{it} = Y_{1,it} + Y_{2,i} + Y_{3,t}$$

- $Y_{2,i}$, $Y_{3,t}$ are the "individual" and "time" specific contribution. These are latent variables. $Y_{1,it}$ is the residual.
- Includes three separate models:
  - single level DSEM: type=general, N=1, $Y_{2,i}$, $Y_{3,t}$ are removed
  - two-level DSEM: type=twolevel, $Y_{3,t}$ is removed
  - cross-classified DSEM: type=cross, full version
- We describe the cross-classified DSEM as it is the most general model, however ....

DSEM model

- Cross-classified DSEM requires that the time scale is aligned for all individuals - not every data set is applicable, ex. observational studies. Time t specific random effect apply for all individuals so time t has to mean the same thing, ex second grade.
- The two-level DSEM much simpler formulation
- The two-level DSEM is the most common and introductory model for applications
- The two-level DSEM can be estimated with less data, fewer requirements for size of N and T as compared to cross-classified DSEM, for example unbalanced designs
- The two-level DSEM easier to estimate as compared to cross-classified DSEM: much fewer number of random effects
- Mplus 8 speed for two-level DSEM always acceptable, Mplus 8 speed for cross-classified DSEM: depends on the model, some models acceptable, models with random variances or random autoregressive parameters can be very slow
- Single level model - one individual modeled separately

DSEM model

- The within level model includes latent variables and observed variables from the previous L (lag) periods

$$Y_{1,it} = \nu_1 + \sum_{l=0}^{L} \Lambda_{1,l}\eta_{1,i,t-l} + \sum_{l=0}^{L} R_l Y_{1,i,t-l} + \sum_{l=0}^{L} K_{1,l} X_{1,i,t-l} + \varepsilon_{1,it}$$

$$\eta_{1,it} = \alpha_1 + \sum_{l=0}^{L} B_{1,l}\eta_{1,i,t-l} + \sum_{l=0}^{L} Q_l Y_{1,i,t-l} + \sum_{l=0}^{L} \Gamma_{1,l} X_{1,i,t-l} + \xi_{1,it}.$$

- Note that all predictors are centered i.e. $Y_{1,i,t-l}$ is not $Y_{i,t-l}$ (covariates X are optional)
- The model is a combination of the state space model with the dynamic factor model, merged with full SEM functionality, e.g., covariates, path analysis, regression among latent variables, CFA, MIMIC, correlated uniqueness, lagged loadings, in addition to the core extensions of two-level and cross-classified modeling as well as categorical variables.

DSEM model

- The usual measurement and structural equations of SEM at level 2 and 3.

$$Y_{2,i} = \nu_2 + \Lambda_2 \eta_{2,i} + \varepsilon_{2,i}$$

$$\eta_{2,i} = \alpha_2 + B_2 \eta_{2,i} + \Gamma_2 x_{2,i} + \xi_{2,i}$$

$$Y_{3,t} = \nu_3 + \Lambda_3 \eta_{3,t} + \varepsilon_{3,t}$$

$$\eta_{3,t} = \alpha_3 + B_3 \eta_{3,t} + \Gamma_3 x_{3,t} + \xi_{3,t}$$

- These include not just between parts of Y_{it} but also observed between level variables

DSEM model

- Random parameters on within level
  - intercepts
  - slopes
  - loadings
  - auto-regressive parameters
  - variances - new V8 feature available for DSEM and non-DSEM
  - random covariance? Only via random factor variances

- We have not found an easy to interpret, random covariance model, that is based on normally distributed random effects which can be used in linear equations as predictors or to be predicted by other variables

DSEM model

- Every within level random parameter s has an individual specific part $s_{2,i}$ and time specific part $s_{3,t}$

$$s = s_{2,i} + s_{3,t}$$

- $s_{2,i}$, $s_{3,t}$ are normally distributed random effects which are a part of the between level latent variable vectors $\eta_{2,i}$ and $\eta_{3,t}$

- Random variances are special

$$s = Exp(s_{2,i} + s_{3,t})$$

- This way we always keep these positive

DSEM model

- The general model on the within level can now also be written with indices i and t for all the possible random parameters

$$Y_{1,it} = \nu_1 + \sum_{l=0}^{L} \Lambda_{1,lit} \eta_{1,i,t-l} + \sum_{l=0}^{L} R_{lit} Y_{1,i,t-l} + \sum_{l=0}^{L} K_{1,lit} X_{1,i,t-l} + \varepsilon_{1,it}$$

$$\eta_{1,it} = \alpha_{1,it} + \sum_{l=0}^{L} B_{1,lit} \eta_{1,i,t-l} + \sum_{l=0}^{L} Q_{lit} Y_{1,i,t-l} + \sum_{l=0}^{L} \Gamma_{1,lit} X_{1,i,t-l} + \xi_{1,it}$$

DSEM model

- The above model assumes conditional normality
- Ordered polytomous and binary dependent variables using the underlying Y^* approach
- Missing data: MAR likelihood based treatment via MCMC estimation. If there is autocorrelation in the data the missing data will be imputed from the neighbouring observations rather than from the average for the person! Note that standard econometrics methodology even for single level models does not include missing data. Even for single level data with missing observations this is new.

Residual DSEM (RDSEM), available in future Mplus release

No change in the between level model. The within level model further splits the autoregressive and the structural part

$$Y_{1,it} = Y_{0,it} + \hat{Y}_{1,it}$$

$$\eta_{1,it} = \eta_{0,it} + \hat{\eta}_{1,it}$$

- The variables $Y_{0,it}$ and $\eta_{0,it}$ represent the linear predictor part (no random element)
- The variables $\hat{Y}_{1,it}$ and $\hat{\eta}_{1,it}$ represent the auto-regressive part and can be thought of as being the residuals

Residual DSEM model

The linear predictor model for $Y_{0,it}$ and $\eta_{0,it}$

$$Y_{0,it} = \nu_1 + \sum_{l=0}^{L} K_{1,lit} X_{1,i,t-l}$$

$$\eta_{0,it} = \alpha_{1,it} + \sum_{l=0}^{L} \Gamma_{1,lit} X_{1,i,t-l}$$

The auto-regressive model for $\hat{Y}_{1,it}$ and $\hat{\eta}_{1,it}$

$$\hat{Y}_{1,it} = \sum_{l=0}^{L} \Lambda_{1,lit} \hat{\eta}_{1,i,t-l} + \sum_{l=0}^{L} R_{lit} \hat{Y}_{1,i,t-l} + \varepsilon_{1,it}$$

$$\hat{\eta}_{1,it} = \sum_{l=0}^{L} B_{1,lit} \hat{\eta}_{1,i,t-l} + \sum_{l=0}^{L} Q_{lit} \hat{Y}_{1,i,t-l} + \xi_{1,it}$$

DSEM Initial Conditions

- At time $t = 1, \ldots, L$ the DSEM model uses predictors with negative time indices such as $\eta_{i,t=0}$, $\eta_{i,t=-1}$, $Y_{1,i,t=0}$, $Y_{1,i,t=-1}$, $X_{i,t=0}$, $X_{i,t=-1}$. We treat these as auxiliary parameters with their own prior.
- If sequences are long such as $T > 50$ the prior does not affect the results. For smaller time-series the priors may have minor effect.
- Mplus implements 2 options
- A. Mplus default: automatic priors, in the first 100 burnin MCMC iterations we update the priors from the sample statistics of η_{it}, $Y_{1,it}$, or $X_{i,t}$, then we discard those 100 MCMC iteration, and retain the constructed priors. Works quite well even for small T.

DSEM Initial Conditions

- B. Specify a normal prior for these auxiliary parameters in model prior. Difficult to use in practice especially when variables are not standardized.

MODEL:	f BY y@1 (&1); f*0.6; y ON f&1*0.4 y&1*0.5 y&2*0.2; y@0.01;
MODEL PRIOR:	f∼N(0,0.6); y∼N(0,1);

DSEM features not available in Mplus 8

- Non-recursive models Y_1 on Y_2, Y_2 on Y_1
- R_0 and Q_0 can not be random
- $\Lambda_{1,l}$, $B_{1,l}$ and random variances, can be random but can not include a time specific random effect
- For categorical variables the lagged variables $Y^*_{i,t-l}$ are not a part of the model. For categorical variables time series models can be built only through latent variables measured by the categorical variable or other continuous dependent or independent variables.

DSEM Estimation

- MCMC with Gibbs sampler. All latent variables, missing values, initial conditions, random effects and model parameters, i.e., all unknown quantities are placed in one of 13 blocks:
  - B1: $Y_{2,i}$
  - B2: All random slopes $s_{2,i}$
  - B3: $Y_{3,t}$
  - B4: All random slopes $s_{3,t}$
  - B5: Other latent variables $\eta_{2,i}$ and $\eta_{3,t}$
  - B6: Latent variables $\eta_{1,it}$, including initial conditions where $t \leq 0$
  - B7: Missing variables Y_{it}
  - B8: Initial conditions $Y_{1,it}$ and $X_{1,it}$ for $t \leq 0$
  - B9: Threshold parameters for all categorical variables θ_3
  - B10: Underlying variables Y^*_{it} for all categorical variables
  - B11: Non-random intercepts, slope and loadings parameters θ_1
  - B12: Non-random variance, covariance and correlation parameters θ_2
  - B13: Random variance parameters

DSEM Estimation

- Determine each block conditional distribution, given all other blocks and the data
- Update (generate new values for) each block from that conditional distribution
- Repeat cycling between the blocks until convergence and use the generated values as the posterior distribution, point estimates, SE
- Mplus mini-max strategy for block formation: minimize the number of block while keeping conditional distributions explicit, i.e., maximizing the blocks. Each block is further split into the sub-blocks that are conditionally independent and update these separately. Strategy for most efficient computation and mixing. Blocks 3,6,7 sequentially updated.
- Bayes estimation inheritance: DSEM algorithm is an extension of Mplus 7.4, i.e., not developed from scratch.
- All conditional distribution are described in the DSEM paper

Overview

- The general DSEM model and estimation
- **Model fit: DIC, model estimated means and variances**
- Centering
- The MEAR model
- Covariates in DSEM, new RDSEM model (residual DSEM)

DIC

- DIC can be used to compare DSEM models. Implemented for models with all continuous dependent variable (no categorical).

$$D(\theta) = -2\log(p(Y|\theta))$$

$$p_D = \bar{D} - D(\bar{\theta})$$

$$DIC = D(\bar{\theta}) + 2p_D$$

- Despite the clear definition with the above formulas, there is substantial variation in what DIC actually is. The source of the variation is the definition of θ, and if it includes the latent variables or not.
- Different definitions of DIC are not comparable. You can compare only if they are using the same likelihood $[Y|\theta]$
- DIC most likely can not be used to compare models if the two models use different θ

DIC

- In DSEM the following are used in the θ vector in addition to all model parameters
  - $Y_{2,i}$ and all random effects $s_{2,i}$
  - $Y_{3,t}$ and all random effects $s_{3,t}$
  - Initial conditions
  - Latent variables $\eta_{1,it}$ if their lagged variables are used
  - Missing variables Y_{it} if their lagged variable is used
- This choice of θ yields easy computation of $[Y|\theta]$ (improved in 8.1)
- To compare two models with DIC all you need to verify is that θ between the two models represents the same random effects. Random effect with zero variance count both ways: fixed or random.
- p_D - estimated number of parameters should generally be near the size of the vector θ, i.e., should be near the count of the above list
- In DSEM p_D is large and needs extra long MCMC sequence for stable estimate

Model fit evaluation based on comparing sample and model estimated statistics

- Assuming stationarity of the autoregressive part of the DSEM model we compute subject specific model estimated mean, variances, autocorrelations of lag L. These can be compared to their sample counterparts.
- Caution about non-stationary models where trend is in the model: use RDSEM style models such as MEAR where trend is separated from the autoregression. The Mplus residual output does not apply directly and no warning is given. This applies only to the residual output in Mplus and not to the model estimation and results. RDSEM and MEAR models, separation of trend and autoregression will be discussed later.

Model fit evaluation based on comparing sample and model estimated statistics

- Model fit evaluation using MSE and correlation between sample v.s. model estimated. For example, means.

$$R = Cor(\mu_i, \overline{Y_{i*}})$$

$$MSE = \sum_{i=1}^{N} (\mu_i - \overline{Y_{i*}})^2 / N.$$

- The correlation is available in the Mplus plot utilities. MSE requires saving the plot data and computing it separately.
- Model estimated auto-correlation is available in the Mplus residual output.
- This method if fully valid only when there is no missing data. In the presence of missing data estimated quantities maybe more accurate than the sample quantities and may not match due to the missing data.

Time-series model estimated means, variance, correlations using Yule-Walker assuming stationarity

$$Z_t = \mu + \sum_{l=1}^{L} A_l Z_{t-l} + \zeta$$

$$\Sigma = Var(\zeta)$$

$$E(Z_t) = \left(I - \sum_{l=1}^{L} A_l \right)^{-1} \mu$$

$$\Gamma_j = Cov(Z_t, Z_{t-j})$$

$$\begin{bmatrix} \Gamma_0 & \Gamma_1^T & \Gamma_2^T & \dots & \Gamma_L^T \\ \Gamma_1 & \Gamma_0 & \Gamma_1^T & \dots & \Gamma_{L-1}^T \\ \Gamma_2 & \Gamma_1 & \Gamma_0 & \dots & \Gamma_{L-2}^T \\ \dots & \dots & \dots & \dots & \dots \\ \Gamma_L & \Gamma_{L-1} & \Gamma_{L-2} & \dots & \Gamma_0 \end{bmatrix} \begin{bmatrix} I \\ -A_1^T \\ -A_2^T \\ \dots \\ -A_L^T \end{bmatrix} = \begin{bmatrix} \Sigma \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

Overview

- The general DSEM model and estimation
- Model fit: DIC, model estimated means and variances
- **Centering**
- The MEAR model
- Covariates in DSEM, new RDSEM model (residual DSEM)

Simulation studies: Centering

- Simulation example using two-level random autoregressive AR(1) model
- Mplus latent centering

$$Y_{it} = \mu_i + r_i(Y_{i,t-1} - \mu_i) + \xi_{it}.$$

- Observed centering

$$Y_{it} = \mu_i + r_i(Y_{i,t-1} - \overline{Y_{i*}}) + \xi_{it}$$

- Uncentered

$$Y_{it} = \mu_i + r_i Y_{i,t-1} + \xi_{it}$$

Centering

- Autocorrelation bias
  - Nickell, S. (1981). Biases in dynamic models with fixed effects. Econometrica: Journal of the Econometric Society, 1417-1426.
  - Hamaker E.L. and Grasman R.P.P.P. (2015) To center or not to center? Investigating inertia with a multilevel autoregressive model. Front. Psychol., 5, 1492.
- Bias for the big-fish-little-pond effect (BFLPE)
  - Ludtke, O., Marsh, H.W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. Psychological Methods,13,203-29.
  - Asparouhov, T. & Muthén, B. (2006). Constructing covariates in multilevel regression. Mplus Web Notes: No. 11.

Centering

- Ludtke / BFLPE bias is for two-level models, involves 2 different variables, and the bias is on the between

$$\frac{(\beta_w - \beta_b)\psi_w}{T\psi_b + \psi_w}$$

- Nickell / autocorrelation bias is for DSEM, involves 1 variable, and the bias is on the within

$$-\frac{1+r}{T-1}$$

- Both stem from not accounting for the error in the sample mean estimate of the mean
- Both disappear when cluster sample size T increases
- Both can appear in parallel in the same example

Centering

- Note that observed centering or uncentered do not exist in case there is missing data - listwise deletion is not an option
- Hamaker and Grasman (2015) show that the uncentered method eliminates Nickell's bias. It does create other bias however, ex for σ_{11}
- Hamaker and Grasman (2015) show that using the true mean to center still creates bias, which means that the bias is not simply a measurement error bias but also the latent nature of the predictor should be accounted for.

Centering - results

Table: Nickell's bias for r=0.3

T	N	Latent centering	Observed centering	Nickell's formula
10	100	0.025	-0.140	-0.144
20	50	0.006	-0.070	-0.068
30	30	0.008	-0.042	-0.045
50	50	0.000	-0.029	-0.027
100	100	-0.001	-0.014	-0.013

Nickell's formula is very accurate. Latent centering eliminates Nickell's bias.

Centering - results

Table: Bias for $Var(\mu_i) = 3$

T	N	latent centering	Uncentered
10	100	-0.015	-1.637
20	50	0.217	-1.483
30	30	0.645	-1.256
50	50	0.378	-1.361
100	100	0.096	-1.508

For latent centering bias on $Var(\mu_i)$ as N increases (or with using weakly informative priors). For the uncentered method in will not disappear even asymptotically as the fundamentals of the model are wrong.

Centering - comparison of latent centering and uncentered

- Mplus latent centering

$$Y_{it} = \mu_i + r_i(Y_{i,t-1} - \mu_i) + \xi_{it}.$$

$$Y_{it} = \mu_i(1 - r_i) + r_i Y_{i,t-1} + \xi_{it}.$$

- Uncentered

$$Y_{it} = \mu_i + r_i Y_{i,t-1} + \xi_{it}$$

- The uncentered and the latent centering are reparameterizations of each other. To obtain the correct μ_i we need to divide by $1 - r_i$
- The latent centering has the advantage of obtaining μ_i directly

Centering - comparison of latent centering and uncentered with subject specific covariate X

- Mplus latent centering

$$Y_{it} = \mu_i + \beta X_i + r_i(Y_{i,t-1} - \mu_i - \beta X_i) + \xi_{it}.$$

$$Y_{it} = \mu_i(1 - r_i) + \beta(1 - r_i)X_i + r_i Y_{i,t-1} + \xi_{it}.$$

- Uncentered

$$Y_{it} = \mu_i + \beta X_i + r_i Y_{i,t-1} + \xi_{it}$$

- The uncentered and the latent centering are NOT reparameterizations of each other as the X_i effect is random in the latent centering.

Overview

- The general DSEM model and estimation
- Model fit: DIC, model estimated means and variances
- Centering
- **The MEAR model**
- Covariates in DSEM, new RDSEM model (residual DSEM)

ARMA(1,1) and the measurement error AR(1) models

- The ARMA(1,1) model

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

$$\sigma = Var(\varepsilon_t)$$

- The measurement error AR(1) = MEAR(1)

$$Y_t = \mu + f_t + \varepsilon_t$$

$$f_t = \phi f_{t-1} + \xi_t$$

$$\sigma_1 = Var(\varepsilon_t), \sigma_2 = Var(\xi_t)$$

- The two models are equivalent

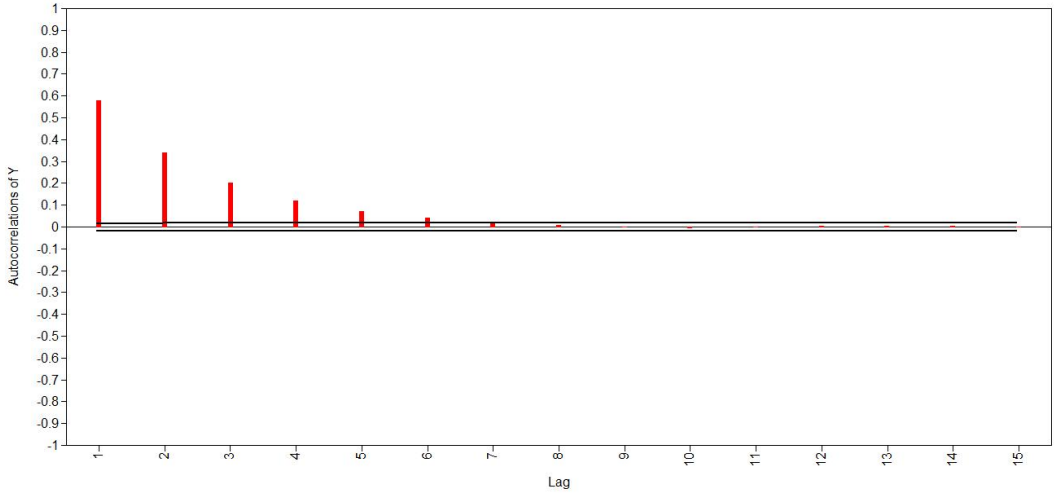
$$\sigma_1 = -\frac{\theta\sigma}{\phi}$$

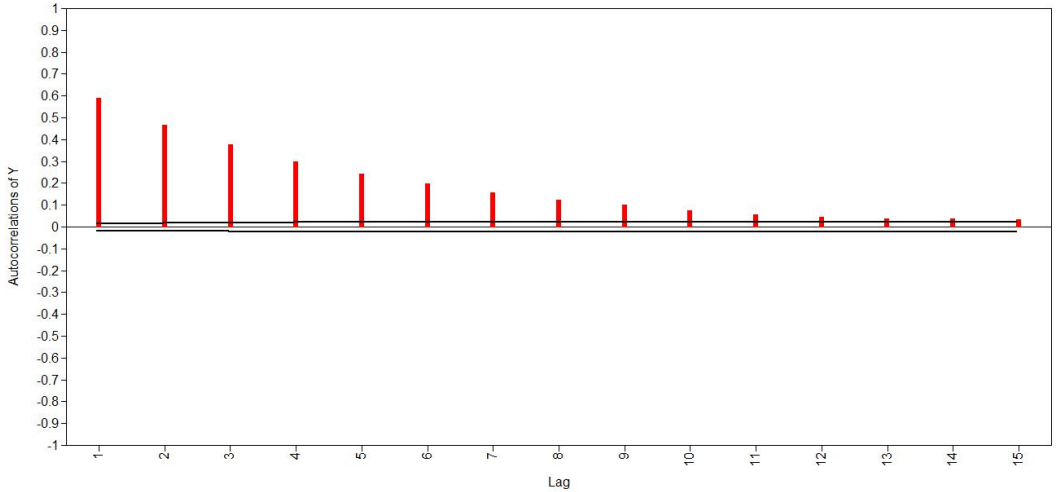
$$\sigma_2 = (1+\theta^2)\sigma + \frac{(1+\phi^2)\theta\sigma}{\phi}$$

ARMA(1,1) and the measurement error AR(1) models

- ... as long as $\sigma_1 > 0$ and $\sigma_2 > 0$
- ϕ is the autocorrelation for f but the autocorrelation for Y is smaller
- The MEAR(1) shows how traditional SEM logic doesn't hold for DSEM: one indicator factor model is perfectly identified
- Two reasons to prefer MEAR(1) v.s. ARMA(1,1)
  - More efficient Mplus estimation
  - Easier to interpret - SEM like flavor
- Two reasons to prefer MEAR(1)/ARMA(1,1) v.s. AR(1)
  - AR(1) exponential decay of autocorrelation is not realistic
  - ARMA(1,1) is a two-parameter fit for the autocorrelation function , v.s., AR(1) which is one parameter
- Easy to test MEAR(1)/ARMA(1,1) v.s. AR(1) using significance of parameter.

AR(1) v.s. ARMA(1,1) autocorrelation decay





ARMA(1,1) simulation study

Table: Bias(coverage) MEAR(1) / ARMA(1,1), N=1

parameter	True value	$T = 100$	$T = 200$	$T = 300$	$T = 500$
μ	0	-.09(.82)	-.01(.89)	-.04(.85)	-.02(.87)
ϕ	.8	-.07(.96)	-.04(.92)	-.03(.87)	-.01(.95)
σ_1	1	-.10(.97)	-.09(.94)	-.08(.88)	-.04(.90)
σ_2	1	.25(.95)	.17(.92)	.14(.91)	.08(.90)

- $T \geq 200$ recommended for small bias and acceptable coverage levels
- For two-level models smaller T are acceptable as long as not all four parameters are subject specific - typically σ_1 and σ_2 will not be subject specific

ARMA(2,1) simulation study

```
title: ARMA(2,1)

montecarlo:
    names = y;
    nobs = 500;
    nreps = 100;
    lagvar=y(2);

analysis: estimator=bayes;

model montecarlo:
    f by y@1 (&1); f*0.6;
    y on f&1*0.4 y&1*0.5 y&2*0.2;
    y@0.00000001;

model:
    f by y@1 (&1); f*0.6;
    y on f&1*0.4 y&1*0.5 y&2*0.2;
    y@0.01;
```

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t + \beta_3 \varepsilon_{t-1}$$

ARMA(2,1) simulation study results

MODEL RESULTS

		ESTIMATES		S. E.	M. S. E.	95%	% Sig
	Population	Average	Std. Dev.	Average		Cover	Coeff
F BY							
Y	1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
Y ON							
F&1	0.400	0.3875	0.2143	0.2429	0.0456	0.960	0.293
Y ON							
Y&1	0.500	0.5178	0.2079	0.2373	0.0431	0.949	0.859
Y&2	0.200	0.1765	0.1754	0.1933	0.0310	0.960	0.111
Intercepts							
Y	0.000	-0.0016	0.0540	0.0466	0.0029	0.869	0.131
Variances							
F	0.600	0.5809	0.0373	0.0392	0.0017	0.939	1.000
Residual Variances							
Y	0.010	0.0100	0.0000	0.0000	0.0000	1.000	0.000

Two-level ARMA(1,1)/MEAR(1) for categorical variables

- Since the underlying variable of a categorical variable is not available for lag modeling MEAR(1) is a good fit and allows time series modeling with a single categorical variable
- For binary variable

$$P(Y_{it} = 1) = \Phi(\mu_i + f_{it})$$

$$f_{it} = \phi f_{i,t-1} + \xi_{it}$$

$$\mu_i \sim N(\mu = -\tau, \sigma_b), \sigma_2 = \sigma_w = Var(\xi_{it})$$

$\sigma_1 = 1$ is the residual variance of Y^*_{it}

- For ordered polytomous variable

$$P(Y_{it} = j) = \Phi(\tau_{j+1} - \mu_i - f_{it}) - \Phi(\tau_j - \mu_i - f_{it})$$

$$f_{it} = \phi f_{i,t-1} + \xi_{it}$$

$$\mu_i \sim N(0, \sigma_b), \sigma_2 = \sigma_w = Var(\xi_{it}), \sigma_1 = 1$$

Two-level ARMA(1,1)/MEAR(1) for binary

Table: Two-level ARMA(1,1)/MEAR(1) with binary variable, N=100, T=300

parameter	True value	Estimate(Coverage)
μ	0	0.00 (.95)
ϕ	.5	0.50(.78)
σ_w	1	1.01(.71)
σ_b	0.5	0.52(.94)

Further algorithmic improvements (8.1) possible to eliminate the need for the MEAR modeling and lower the requirements for the size of T (also not needed if you have more than one factor indicator).

Two-level ARMA(1,1)/MEAR(1) for ordered polytomous

Table: Two-level ARMA(1,1)/MEAR(1) with ordered polytomous, N=100, T=100

parameter	True value	Estimate(Coverage)
τ_1	-3	-3.06 (.87)
τ_2	-1	-1.02 (.81)
τ_3	0	-0.01 (.79)
τ_4	1	1.01 (.75)
τ_5	3	3.05 (.81)
ϕ	.5	0.50(.93)
σ_w	1	1.09(.83)
σ_b	0.5	0.54(.94)

Overview

- The general DSEM model and estimation
- Model fit: DIC, model estimated means and variances
- Centering
- The MEAR model
- **Covariates in DSEM, new RDSEM model (residual DSEM)**

How to add a covariate in ARMA(1,1) and AR(1) models

- The same applies for AR(1) and ARMA(1,1)/MEAR(1). Three ways to do it (using N=1 as an example)
- Residual AR model

$$Y_t = \mu + f_t + \beta_1 X_t + \xi_t$$

$$f_t = \phi f_{t-1} + \varepsilon_t.$$

- Full AR model

$$Y_t = \mu + f_t + \xi_t$$

$$f_t = \phi f_{t-1} + \beta_2 X_t + \varepsilon_t.$$

- Joint effect model

$$Y_t = \mu + f_t + \beta_1 X_t + \xi_t$$

$$f_t = \phi f_{t-1} + \beta_2 X_t + \varepsilon_t.$$

How to add a covariate in ARMA(1,1) and AR(1) models

- Consider the fundamental difference between the models by what it implies for $E(Y_t|X)$
- Residual AR model - no effect beyond the last value of X

$$E(Y_t|X) = \mu + \beta_1 X_t.$$

- Full AR model - accumulation effect of X with diminishing effects

$$E(Y_t|X) = \mu + \beta_2(X_t + \phi X_{t-1} + \phi^2 X_{t-2} + \phi^3 X_{t-3} + ...).$$

- Joint effect model - direct and accumulated effect

$$E(Y_t|X) = \mu + \beta_1 X_t + \beta_2(X_t + \phi X_{t-1} + \phi^2 X_{t-2} + \phi^3 X_{t-3} + ...).$$

How to add a covariate in ARMA(1,1) and AR(1) models. The special case of $X_t = t$, linear growth model

- Hamaker, E.L. (2005) Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. Sociological Methods and Research, 33, 3, 404 - 418.
- It is shown in this paper that the residual AR and the full AR models are equivalent, i.e., the joint effect model is not identified

How to add a covariate in ARMA(1,1) and AR(1) models. The special case of $X_t = t$, linear growth model

- The residual AR model: linear growth AR(1) model

$$Y_t = \gamma_0 + \gamma_1 t + \xi_t$$
$$\xi_t = \phi \xi_{t-1} + \varepsilon_t$$

- The full AR model: linear growth full AR(1) model

$$Y_t = \beta_0 + \beta_1 t + \phi Y_{t-1} + \varepsilon_t$$

- Here t affects Y_t through its effect on Y_{t-1} in addition to the direct effect β_1

$$Y_{t-1} = \beta_0 + \beta_1 (t-1) + \phi Y_{t-2} + \varepsilon_{t-1}$$

- The two models are algebraically equivalent and the joint effect model is unidentified

$$\gamma_0 = \frac{\beta_0}{1-\phi} - \frac{\phi \beta_1}{(1-\phi)^2}, \quad \gamma_1 = \frac{\beta_1}{1-\phi}$$

AR(1) quadratic growth model, $X_t = (t, t^2)$

- The quadratic growth AR(1) model

$$Y_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \xi_t$$
$$\xi_t = \phi \xi_{t-1} + \varepsilon_t$$

- The quadratic growth full AR(1) model

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \phi Y_{t-1} + \varepsilon_t$$

- The two models are algebraically equivalent and the joint effect model is unidentified. No simple reparametrization.

$$\gamma_0 = \frac{\beta_0}{1-\phi} - \frac{\phi\beta_1}{(1-\phi)^2} + \frac{\beta_2\phi(1+\phi)}{(1-\phi)^3}$$
$$\gamma_1 = \frac{\beta_1}{1-\phi} - \frac{2\phi\beta_2}{(1-\phi)^2}$$
$$\gamma_2 = \frac{\beta_2}{1-\phi}$$

Two-level joint effect ARMA(1,1)/MEAR(1) simulation study

$$Y_{it} = \mu_i + f_t + \beta_1 X_{it} + \xi_{it}$$

$$f_{it} = \phi f_{i,t-1} + \beta_2 X_{it} + \varepsilon_{it}$$

$$\mu_i \sim N(\mu, \sigma_b)$$

- $\beta_1 = 0.3$, $\beta_2 = 0.4$, $\phi = 0.5$, $\mu = 0$, $\sigma_b = 0.7$, $Var(\xi_{it}) = Var(\varepsilon_{it}) = 1$
- The covariate is generated using AR(1) process with $Var(X_{it}) = 1$ and autocorrelation $r_x = 0, 0.5, 0.8$
- We analyze the data using the joint effect model, the residual AR, and the full AR

Two-level joint effect ARMA(1,1) simulation results: joint effect

Table: Two-level joint effect ARMA(1,1) with covariate, N=200, T=100

parameter	r_x	True value	Estimate(Coverage)
β_1	0	.30	.30(.87)
β_1	0.5	.30	.30(.96)
β_1	0.8	.30	.31(.89)
β_2	0	.40	.40(.87)
β_2	0.5	.40	.40(.93)
β_2	0.8	.40	.40(.90)
ϕ	0	.50	.50(.88)
ϕ	0.5	.50	.50(.93)
ϕ	0.8	.50	.50(.93)

No bias. Good coverage. Model is well identified.

Two-level joint effect ARMA(1,1) simulation results: residual AR

Table: Two-level joint effect ARMA(1,1) with covariate analyzed as residual AR, N=200, T=100, dropping β_2

parameter	r_x	True value	Estimate(Coverage)
β_1	0	.70	.65(.00)
β_1	0.5	.70	.74(.07)
β_1	0.8	.70	.88(.00)
ϕ	0	.50	.50(.92)
ϕ	0.5	.50	.51(.85)
ϕ	0.8	.50	.52(.83)

Both parameters are biased. Coverage is low. Bias depends on r_x.

Two-level joint effect ARMA(1,1) simulation results: full AR

Table: Two-level joint effect ARMA(1,1) with covariate analyzed as full AR, N=200, T=100, dropping β_1

parameter	r_x	True value	Estimate(Coverage)
β_2	0	.70	.69(.92)
β_2	0.5	.70	.67(.21)
β_2	0.8	.70	.65(.07)
ϕ	0	.50	.36(.00)
ϕ	0.5	.50	.38(.00)
ϕ	0.8	.50	.41(.00)

Both parameters are biased. Coverage is low. Bias depends on r_x.

How to add a covariate in ARMA(1,1) and AR(1) models

- The residual AR and the full AR models are not a reparameterization of each other.
- The choice between the residual AR and the full AR models should not be made based on what is simpler to interpret, software availability, or tradition - rather the data should decide that.
- For the AR(1) model the issue is more complicated than the ARMA(1,1) model as the residual AR model requires fixing $Var(\xi_i)$ to 0 (small value near zero) which leads to slow convergence, i.e., we have to specify the model as MEAR(1)
- For the AR(1) model the Residual DSEM is the appropriate approach - fast convergence, no fixing residuals to zero, no MEAR(1) usage. RDSEM absorbs the complexity but is complicated to estimate than DSEM.

Simulation study: two-level residual AR(1) model

$$T = 50, N = 50, rep = 50, X_{it} \sim AR(1), r_x = 0.5$$

$$Y_{it} = \mu_i + \beta X_{it} + \xi_{it}$$

$$\xi_{it} = \phi \xi_{it-1} + \varepsilon_{it}$$

- Three methods: DSEM-MEAR(1) with fixed residual variance to a small value, DSEM with free residual variance, RDSEM (new input)

```
MODEL:

%WITHIN%
y on x*1; y*1;
y^ on y^1*0.7;

%BETWEEN%
y*1; [y*3];
```

Simulation study: two-level residual AR(1) model

Table: Two-level residual AR(1) with covariate bias(coverage)

method	DSEM-fixed	DSEM-free	RDSEM
β	.00(.60)	.00(.55)	.00(.90)
ϕ	.03(.62)	.01(.92)	.00(.94)
time per rep in sec	3	33	0.5

Clearly RDSEM outperforms DSEM for this model due to quality of mixing - only 200 MCMC iterations until convergence.

Linear growth AR(1) simulation

In this example we illustrate the following 8 concepts

- Equivalence of residual and full AR(1) models for linear growth models
- We illustrate that the model estimation is correct even when stationarity is not present in the model and that only residual/tech4/stand depend on the stationarity assumption
- The dependence of Yule-Walker output: residual/tech4/stand on the stationarity of the autoregressive portion of the model
- How to setup the model so that the trend is not included in the autoregressive portion of the model using the MEAR concept for the purpose of obtaining correct Yule-Walker output
- We illustrate the advantages of RDSEM
- Challenges with fixing variance to zero/small values
- We illustrate 4 ways to run the model in Mplus and obtain the exact same model estimates
- How to use MODEL CONSTRAINT to implement the reparameterization for the residual and full growth AR(1) models

Linear growth AR(1) simulation

We generate data 500 points using the full AR linear growth model

$$Y_t = \beta_0 + \beta_1 t + \phi Y_{t-1} + \varepsilon_t$$

$\beta_0 = 1$, $\beta_1 = 0.3$, $\phi = 0.5$, $\theta = Var(\varepsilon_t) = 1$, $t = 0.1, 0.2, 0.3, \ldots, 50$.

Linear growth AR(1) simulation

Estimate the following 4 equivalent models

- Linear growth full AR(1), followed by model parameter transformation, using DSEM

$$Y_t = \beta_0 + \beta_1 t + \phi Y_{t-1} + \varepsilon_t$$

- Linear growth residual AR(1) using RDSEM

$$Y_t = \gamma_0 + \gamma_1 t + \varepsilon_t$$

$$\varepsilon_t = \phi \varepsilon_{t-1} + \xi_t$$

- Linear growth MEAR(1) model with $Var(\zeta_t)$ fixed to a small value 0.05, DSEM

$$Y_t = \gamma_0 + \gamma_1 t + f_t + \varepsilon_t$$

$$f_t = \phi f_{t-1} + \xi_t$$

- Linear growth MEAR(1) model with $Var(\zeta_t)$ estimated as a free parameter, DSEM

Linear growth full AR(1) using DSEM

```
variable:   names=y t;
            lagged=y(1);
data:       file=1.dat;

analysis:   estimator=bayes;
            proc=2; fbiter=50000;

model:
 y on y&1 (phi)
       t  (b1);
 [y] (b0);

model constraint: new(g0 g1);
 g0=b0/(1-phi)-0.1*b1*phi/(1-phi)**2;
 g1=b1/(1-phi);

output: residual;
```

Linear growth residual AR(1) using RDSEM

```
variable: names = y t;
          lagged=y(1);

data: file=1.dat;

analysis: estimator=bayes;
          proc=2; fbiter=50000;

model:
 y on t*0.6;
 y^ on y^1;
 [y*0];

output: residual;
```

Linear growth MEAR(1) (free) using DSEM

```
variable:  NAMES ARE y t;

data: file=1.dat;

analysis:  estimator=bayes;
           proc=2; fbiter=50000; thin=50;

model:
 y on t;
 f on f&1*0.5 ;
 f by y (&1);
 y*0.05;
 [y*0];

output: residual;
```

Linear growth AR(1) simulation: results

	True value	DSEM Full AR	RDSEM Resid AR	MEAR(1) free Full AR	MEAR(1) fixed Full AR
convergence		fast	fast	slow	slow
γ_0	1.94	1.94(.18)	1.94(.19)	1.93(.19)	1.93(.19)
γ_1	.6	.6(.006)	.6(.006)	.6(.006)	.6(.006)
ϕ	.5	.51(.04)	.51(.04)	.53(.04)	.53(.04)
θ	1	1(.06)	1(.06)	.93 (.09)	.94(.06)
$E(Y)$	16.94	23.33	16.94	16.94	16.94
$Var(Y)$	76.23	67.06	76.05	76.12	76.17
DIC		1421	1420	N/A	N/A

Linear growth AR(1) simulation: conclusions

- All four methods produce the same model estimates, MEAR models produce approximation
- Full AR(1) DSEM model yields incorrect residual / tech4 / stand results due to violating the assumptions of Yule-Walker
- MEAR(1) free is preferred over MEAR(1) fixed because you don't need to decide on the small value, also safeguard against the small variance not being zero, i.e., safeguard against the model not being AR(1) but rather ARMA(1,1)
- MEAR(1) formulation shows slow convergence, however, in real data sets most likely the residual variance will not be exactly zero / much better convergence
- MEAR(1) methods do not produce usable DIC due to conditioning on the within level factor $Var(Y|f) \approx 0$
- Fixing variances to zero is not harmless in Bayes estimation as it is in ML - most likely causing poor mixing and convergence
- RDSEM is the best method - not available in Mplus 8, available in Mplus 8.1