Dynamic Structural Equation Models

Tihomir Asparouhov and Bengt Muthén

January 13, 2016

Overview

- Motivation
- General DSEM framework and estimation
- DIC
- Empirical example
- Simulation Studies: DAFS, WNFS, 3Level AR(1)
- Unevenly spaced and individual-specific times of observations
- Dynamic Latent Class Analysis: Hidden Markov Chains in Mplus - long LTA (latent transition analysis)
- M3 pre-conference workshop on Mplus Version 8, May 23, 2016 at the University of Connecticut

Motivation

- Merge ”time series” and ”structural equation” modeling concepts in a generalized modeling framework in Mplus V8
- Until recently, most dynamic structural equation models were focused on the case N=1, due to connection with econometrics, ARMA models and Kalman filer estimation. Most social science and biostatistics/epidemiological applications have $N > 1$. Thus time-series SEM model must be a two-level model where the cluster variable is the individual and we have a long sequence of observations for that individual.
- Two distinct sources of correlation: within individual correlations can be due to subject-specific effect (two-level modeling) or it can be due to correlation due to proximity (autocorrelation) of the observations (time series modeling). The two types of correlations are easy to parse out from the data in sufficiently long longitudinal data.

Motivation continued

- The need for this modeling arises if we want to see continuum of latent variable constructs and from the fact that we don't know how often to measure the latent construct. If we measure too often we have large autocorrelation and false independence, if we measure too infrequently we can miss important changes and portray incomplete picture
- Ignoring the autocorrelation will produce incorrect SE and hide a substantial amount of the underlying relationship in the data.
- When the autocorrelation is substantial and we don't account for it, we essentially ignore the best predictors - the values from the previous time period
- Current modeling approach to long longitudinal data using ML estimation: A. Two-level modeling (ignoring proximity correlation and assuming observations are independent conditional on the individual) B. Multivariate model (computationally intractable beyond time length of 100 observations)

Motivation continued

- ML is not a feasible estimation method for two-level time-series model, and is fairly complex even for single level.
- Bayesian estimation is feasible. Adding ”one additional correlation” results in changing one step in the MCMC

Motivation continued

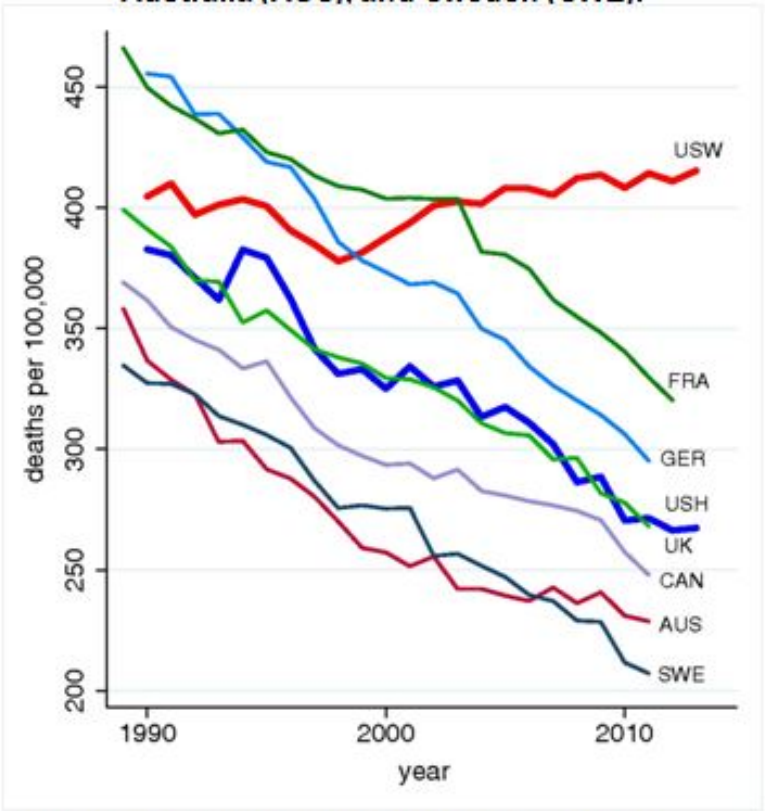
- Increasing demand for methods analyzing daily or weekly data. Very long longitudinal models. Frequent observations leading to substantial correlation in consecutive observations.
- Various mobile devices utilized for collection of time series measurements: ecological momentary assessment (EMA) data and experience sampling methods (ESM). Increased popularity of time varying effect models (TVEM).
- Continuous time monitoring of latent constructs: such as depression, motivation, math aptitude, smoking urge.
- Vast potential for new application, new improved solutions for existing problems. Potential to improve health care via real-time mobile self-monitoring, early detection critical condition, for overdosing, drug interactions, opiates abuse, alcohol abuse, depression monitoring, suicide prevention etc.
- Health care accountability - various laws are passed in US requiring health care providers to collect data, show progress and effective treatment.

Motivation continued

- Why do we need continuous monitoring of latent constructs. We could detect problems earlier and prevent/react.
- Mortality in US is increasing. Life expectancy is decreasing. It took 25 years of data? How long would it take if we used weekly or monthly data.
- Imagine a future where mobile self monitoring communicated to your doctor or a computer algorithm can be a trigger for hospitalization to prevent suicide, drug overdose, etc
- We can make the prediction that in 20 years from now there will be more DSEM applications than SEM. Why? Because this is the data we will get and because it is easy. DSEM is NOT more difficult than SEM

Motivation continued

All-cause mortality, ages 45–54 for US White non-Hispanics (USW), US Hispanics (USH), and six comparison countries: France (FRA), Germany (GER), the United Kingdom (UK), Canada (CAN), Australia (AUS), and Sweden (SWE).



Anne Case, and Angus Deaton PNAS 2015;112:15078-15083

©2015 by National Academy of Sciences

PNAS

Mplus general DSEM framework

- Let Y_{it} be an observed vector of measurements for individual i at time t.
- Let η_{it} be a latent variable vector for individual i at time t.
- Let X_{it} be a observed vector of covariates individual i at time t.
- Similarly Y_i, X_i and η_i are individual-specific variables
- Similarly Y_t, X_t and η_t are time specific variables
- Main decomposition equation

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

DSEM framework continued

- The within level model includes previous periods latent variables - modeling the correlations in consecutive periods.
- Only as predictors, not as dependent variables (if we include it as predictor we will have a repeating equation)
- Prior latent variables can affect current latent variables DAFS (direct autoregressive factor score) or current observed variables directly WNFS (white noise factor score)
- One generalized framework to include many models
- L is the lag variable: how many time period back are included in the model. L=1 mean that just the latent variables in the previous period need to be included.

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

$$\eta_{1,i,t} = \alpha_1 + \sum_{l=0}^{L} B_{1,l} \eta_{1,i,t-l} + \Gamma_1 x_{1,it} + \xi_{1,it}.$$

DSEM framework continued

- The usual structural equations at level 2 and 3.

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

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

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

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

- Random intercepts, loadings, slopes, variances: every within level parameter can be random. s_1 non-random within level parameter, $s_{2,i}$ is individual random effect part of $\eta_{2,i}$ and $s_{3,t}$ is time specific random effect part of $\eta_{3,t}$

$$s = s_1 + s_{2,i} + s_{3,t}$$

Behind the scenes DSEM

- There are 3 models in the above specification
- TYPE=CROSSCLASSIFIED; CLUSTER=ID TIME. Modeling time specific effects requires the time scale to be comparable across individuals. No examples in this talk. Long MTMM.
- TYPE=TWOLEVEL; CLUSTER=ID. No time specific random effects (non-random effects are ok). This will be the most common and introductory model to be used.
- TYPE=GENERAL. Single level. N=1 case. No random effects. All observations are correlated. Multivariate econometrics. Kalman filter modeling. ARMA. It can also be used as a wide $N > 1$ multivariate spread.

DSEM extensions

- The above model assumes normality
- Ordered polytomous and Binary dependent variables using the underlying Y^* approach
- Missing data: MAR likelihood based treatment via MCMC imputation
- The above model extends the Version 7 model time-intensive cross-classified SEM model with the time-series capabilities
- Missing data imputation using DSEM - much more realistic when there is autocorrelation in the data as the missing data will be imputed from the neighbouring observations rather than from the average for the person
- MCMC based distributions for all model parameters, latent variables, and individual-specific or time specific random effects

DSEM language

- ETA&1 refers to ETA with lag 1, ETA&2 refers to ETA with lag 2, etc ...
- ETA by Y1-Y5 (& L); specifies the lag L and that ETA&1, ETA&2, ..., ETA&L will be in the model
- DAFS model: ETA by Y1-Y5 (& 1); ETA on ETA&1;
- WNFS model: ETA by Y1-Y5 (& 1); Y1-Y5 on ETA&1;
- What if you want to use Y&1 or X&1 as the predictor? That you can do even now, no need for new algorithms
- ARMA(1,1) model: ETA by Y (& 1); Y on ETA&1 Y_1; Y@0;

$$Y_t = \mu + \beta_1 Y_{t-1} + \varepsilon_t + \beta_2 \varepsilon_{t-1}$$

DSEM Stationarity

- The models as described assumes that the distribution of $Y_t|X_t$ stabilizes i.e. is stationary. If the observed Y_t data does not exhibit a stationary distribution and the distribution changes with time then covariates X_t that change over time need to be included (for example including T as covariate will model a distribution that has a mean changing linearly over time)
- Alternative method for dealing with non-stationarity is to use ARIMA methodology. Instead of modeling Y_t, model $Y_t - Y_{t-1}$ or more generally $(1-L)^d Y_t$ as a stationary process
- How to check that estimated model is stationary? MA (moving average) has no effect - always stationary. The WNFS is always stationary. AR (autoregressive) portion should have all root of the lag-equation > 1, i.e., for AR(1) $Y_t = \beta Y_{t-1} + \varepsilon$, $|\beta| < 1$ for stationarity and it is the autocorrelation. In multivariate form the eigenvalues of $|\beta|$ must be less than 1 by absolute value and $\beta = V^{-1}C$ is not the autocorrelation coefficients and could be bigger than 1.

DSEM Stationarity continued

- How to really check if a model is stationary: generate data in Mplus and see if it explodes
- Consider the simple DAFS model

$$Y_{it} = \mu_i + \Lambda\eta_{it} + \varepsilon_{it}$$

$$\eta_{it} = \alpha + \beta\eta_{i,t-1} + \xi_{it}$$

- This model is stationary if and only if $|\beta| < 1$. If there are more than one factors and η_{it} is a vector then β is a matrix and the process is stationary if and only if the eigenvalues of the matrix are smaller than 1 by absolute value.
- Note also that in the univariate case

$$Var(\eta_{it}) = Var(\xi_{it})/(1-\beta^2), E(\eta_{it}) = \alpha/(1-\beta)$$

- If β is close to 1 then the mean and the variance will be close to infinity

DSEM Estimation

- Four estimating issues as an addition to the Mplus Version 7 algorithm.
- 1 [Between Parts | observed data] can't really be computed as the likelihood is messy and the observations are not independent given the between level random effects $=>$ η has to become part of MCMC and the between level random effects are now [Between Parts | observed data, η]. Once this is done we can use two-level FIML formulas. An efficiency gain in mixing quality that is being lost.
- 2. $[\eta_{ti}|*]$ used in the MCMC estimation now uses $2*L+1$ measurement equations and one regression equation and depends on

$$Y_{t,i}, Y_{t+1,i}, \ldots, Y_{t+L,i},$$

$$X_{t,i}, X_{t+1,i}, \ldots, X_{t+L,i},$$

$$\eta_{t-L,i}, \ldots, \eta_{t+L,i}$$

DSEM Estimation Continued

- 3. Initial Conditions. Now we have $\eta_{i,t=0}$, $\eta_{i,t=-1}$, ... in the model. We treat these as auxiliary parameters with their own prior. If sequences are long such as $T > 50$ it doesn't matter. For smaller sequences it does. Mplus offers 3 options
  - fixed them to zero using prior N(0,0) (currently our default) or
  - specify a non-zero prior or
  - use our implementation of auto priors, in the first 100 MCMC iterations we update the priors from the sample statistics of η, then we discard those 100 MCMC iterations.
- 4. The posterior distribution of η_{ti} changes across t. It is the same in the middle of the sequence but changes at the end or the beginning as there are fewer observation (the sequence is truncated).

DSEM Estimation Continued

- Note also that η with negative indices become important in simulation studies as well. One solution is to draw from the prior of the auxiliary parameters. Second solution is to draw and discard the first 20 draws. After that the simulation distribution stabilizes and is independent of these initial values.
- In Version 8 restriction that is likely to stay: random loadings can vary across individual but not across time. Random slopes however can vary across individual and time with current implementation.

DSEM Estimation Continued

- Do not be surprised or alarmed if Mplus converges in 20 seconds and WinBugs takes 24 hours.
- Here are the four main reasons:
- 1. Mplus uses Fortran (fastest computational environment)
- 2. Mplus uses parallel computing so each chain is computed separately (the number of chains will not increase the computing time as long as you have enough CPUs on your computer (most come with 8 - Quad Core i7 processors). Use the PROC=number_chains command
- 3. Mplus uses the largest updating blocks possible - complicated to program but gives the best mixing quality
- 4. Mplus uses sufficient statistics

DIC

- DIC is currently the only information criterion implemented in Mplus for comparing DSEM models.

$$D(\theta) = -2\log(p(Y|\theta))$$
$$p_D = \bar{D} - D(\bar{\theta})$$
$$DIC = D(\bar{\theta}) + 2p_D$$

- Despite the clear definition with the above formulas, there is substantial variation in what DIC actually is. The sources of the variation come from - what exactly Y is and what exactly θ is. For example in a factor analysis model, one way to compute DIC is using the observed Y implied variance covariance matrix, and never really involve the factor, i.e., likelihood = [Y]. You can also treat the factor as parameters in that case likelihood = [Y|eta]. Also different treatment of missing values: parameters or not a part of the likelihood.

DIC Continued

- Different definitions of DIC are not comparable. You can compare only if they are using the same likelihood definition.
- Software implementations not clear about what definition they use, even published peer reviewed articles limited to the above definition, but no full details on what likelihood is being used
- Different definitions of DIC have different power to detect significance.
- An easy way to figure out what DIC is being computed: look at p_D - estimated number of parameters
- Among the different definitions of DIC: the smaller p_D the better. Fewer parameters to be integrated, faster convergence, more accurate estimate, more power, if p_D is large then it is not usual to have a situation when model estimation has converged but DIC estimation has not

DIC Continued

- Looking at p_D is surprisingly useful, you can detect estimation problems with it. If p_D is not close enough to what you expect it to be most likely you need to run many more MCMC iterations. If p_D is negative you most certainly should not consider the model as converged.
- p_D is the "estimated number of parameters" it does not match the actual number of parameters. Consider random effect with zero variance, this is equivalent to a non-random parameter. p_D has a way of telling this and will not consider these random effects to be different parameters but one parameter. Similarly highly correlated random effects will count as one not two random effects. The more correlated random effects are the more discounted they are. For example two random effects with correlation 0.75 could count as 1.5 parameters. If the correlation is 0 they will count as 2 parameters, if the correlation is 0.95 they will count as 1.2 parameters. All these numbers are approximate.

DIC for DSEM

- All latent variables are conditioned on, between and within level latent variables. Likelihood is now straightforward to compute as the within cluster variables are independent. Large number of parameters.
- In some models this definition becomes useless. Consider a model like ARMA

$$Y_t = \beta_1 Y_{t-1} + \varepsilon_t + \beta_2 \varepsilon_{t-1}$$

The likelihood used by DIC is $[Y_t | Y_{t-1}, \varepsilon_t, \varepsilon_{t-1}]$ which has zero conditional variance and - infinity as likelihood. If your model requires the specification Y@0; on the within level DIC will likely be useless and you should be able to see that in the large negative DIC value. p_D still ok.

DIC for DSEM

- You can use DIC to compare complex hypothesis such as a whole set of parameters is zero, or a set of random effects is zero. You can use DIC to compare the need for random loadings (fixed v.s. random loadings). You can use DIC to compare non-nested models. If testing just one parameter instead use the confidence/credibility interval.

How to use observed variables lags Y&1 and X&1?

- Three ways
- 1. You can enter the lag variable as a separate column in your data
- 2. You can create a latent variable behind the observed variable etaY by Y@1 (& 1); Y@0;
- 3. Use new Mplus data manipulation facility: DATA LAG

Empirical example: borderline personality disorder data

- The example without AR is from Asparouhov and Muthen (2015), ”General Random Effect Latent Variable Modeling: Random Subjects, Items, Contexts, and Parameters”, in Advances in Multilevel Modeling for Educational Research. The data comes from Jahng, Wood, Trull (2008)
- 21 items measuring mood for 84 subjects, two groups according to DSM-IV classification.
- 76 to 186 observations per person all within 4 weeks: violation of equally spaced times.
- Primary interest is to determine if the group variable increases the factor variance: β = SIGMA ON GROUP
- All Version 7 features apply to DSEM: random loadings, random indicator mean, random factor mean, random factor variance

Input file for empirical example

```
model:
%within%
s1-s21 | f by jittery-scornful (& 1);
f@1; rho | f on f&1;

%between%
rho; rho on group;
f; f on group;
s1-s21 jittery-scornful;
[s1-s21*1] (m1-m21);
sigma by s1-s21*1 (m1-m21);
sigma; sigma on group;
```

Results from empirical example: borderline personality disorder data

Model	DIC	p_D	p	β
Full DSEM model	436628	13473	113	.46[.11,.92]
DSEM with non-random AR	436749	13553	111	.24[-.04,.58]
DSEM without AR	437276	14087	110	.29 [-.01,.67]
Twolevel model	460351	3366	110	.29[.04,.65]

Conclusions from empirical example

- Full DSEM model gets the best DIC.
- Dynamic modeling affects point estimates, standard errors, significance
- The Full DSEM model separates mood variation as momentary and global. Fast changing mood (small ar) v.s. slow changing mood (large ar).
- Sine curve amplitude effect: Sigma on Group
- Sine curve frequency effect: Rho on Group
- DSEM-DIC not on the same scale as Twolevel-DIC

Alternative individual-specific variances for a variable

- Sigma | E1 by E; E1@1; E@0
- E=Sigma*E1, where E1 is standard normal
- Sigma is between level random effect
- Var(E | ID=i)=$Sigma_i^2$

Simulation Study - DAFS Lag 3 model

```
montecarlo:
     names = y1-y5;
     NOBS = 10000;
     NREP = 100;
     NCSIZES = 1;
     CSIZES = 100(100);

ANALYSIS:    TYPE IS TWOLEVEL;
             estimator=bayes;
             biter=(500); proc=2;

model montecarlo:
     %within%
     f by y1-y5*1 (& 3);
     y1-y5*1;
     f@1;
     f on f&1*0.4 f&2*0.2 f&3*0.1;

     %between%
     fb by y1-y5*0.5; fb@1; y1-y5*0.2;
```

Simulation Study Results - DAFS Lag 3 model

F	BY							
Y1		1.000	0.9992	0.0134	0.0120	0.0002	0.900	1.000
Y2		1.000	1.0010	0.0127	0.0120	0.0002	0.950	1.000
Y3		1.000	1.0001	0.0138	0.0120	0.0002	0.910	1.000
Y4		1.000	0.9997	0.0128	0.0120	0.0002	0.930	1.000
Y5		1.000	0.9988	0.0134	0.0119	0.0002	0.910	1.000
F	ON							
F&1		0.400	0.4024	0.0121	0.0123	0.0002	0.980	1.000
F&2		0.200	0.1983	0.0143	0.0139	0.0002	0.950	1.000
F&3		0.100	0.0997	0.0138	0.0126	0.0002	0.910	1.000

Simulation Study - WNFS Lag 5 model

```
montecarlo:
    names = y1-y5;
    NOBS = 10000;
    NREP = 100;
    NCSIZES = 1;
    CSIZES = 100(100);

ANALYSIS:   TYPE IS TWOLEVEL;
            estimator=bayes;
            biter=(500); proc=2;

model montecarlo:
    %within%
    f by y1-y5*1 (& 5);
    y1-y5*1;
    f@1;
    y1-y5 on f&1*0.4 f&2*0.2 f&3*0.3 f&4*0.2 f&5*0.1;

    %between%
    fb by y1-y5*0.5; fb@1; y1-y5*0.2;
```

Simulation Study Results - WNFS Lag 5 model

Y1	ON							
F&1		0.400	0.4012	0.0155	0.0160	0.0002	0.960	1.000
F&2		0.200	0.1989	0.0168	0.0161	0.0003	0.940	1.000
F&3		0.300	0.3021	0.0145	0.0165	0.0002	0.980	1.000
F&4		0.200	0.1998	0.0179	0.0167	0.0003	0.950	1.000
F&5		0.100	0.0985	0.0155	0.0164	0.0002	0.960	1.000

Simulation Study - DAFS-WNFS Combo Lag 1 model - ARMA(1,1) factor

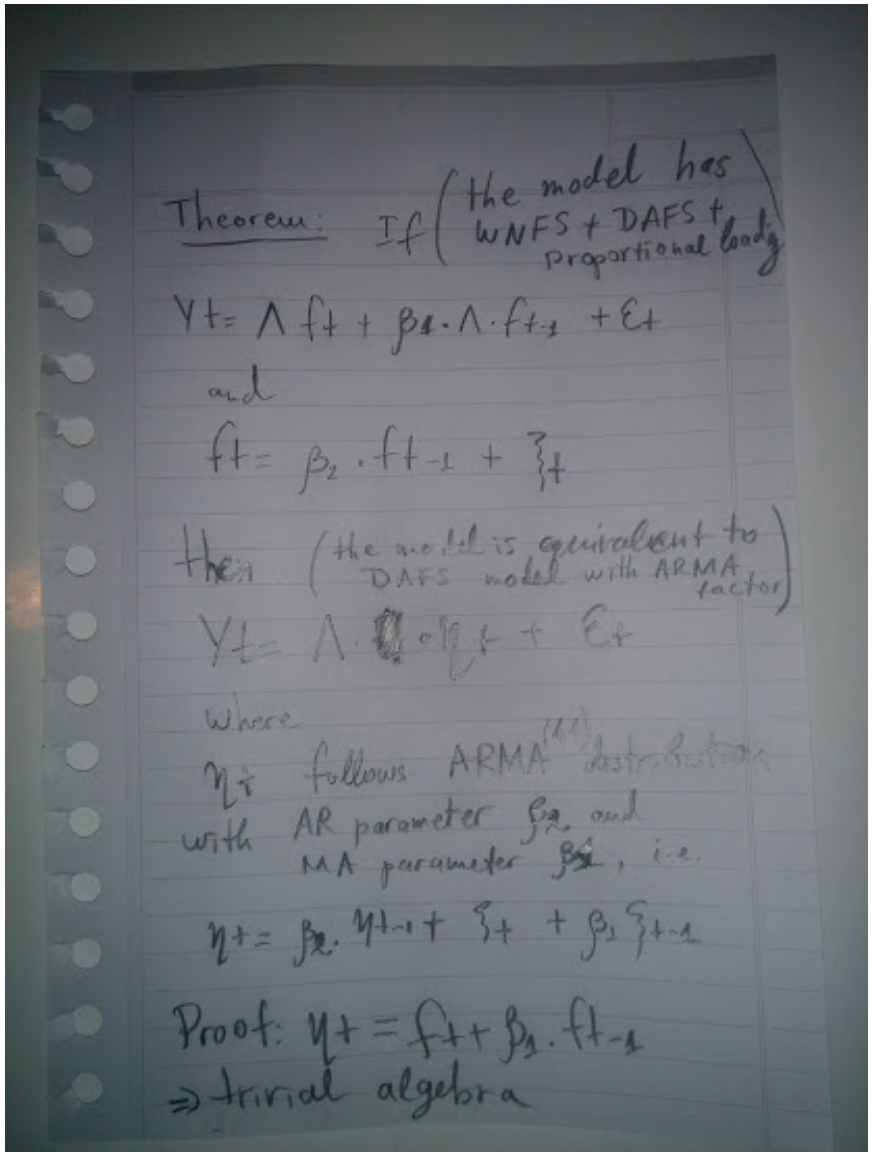
```
montecarlo:
    names = y1-y5;
    NOBS = 10000;
    NREP = 100;
    NCSIZES = 1;
    CSIZES = 100(100);

ANALYSIS:   TYPE IS TWOLEVEL;
            estimator=bayes;
            biter=(500); proc=2;

model montecarlo:
    %within%
    f by y1-y5*1 (& 1);
    y1-y5*1;
    f@1;
    f on f&1*0.4;
    y1-y5 on f&1*0.6;

    %between%
    fb by y1-y5*0.5; fb@1; y1-y5*0.2;
```

Simulation Study Results - DAFS-WNFS Combo Lag 1 model - ARMA(1,1) factor

F	ON							
F&1		0.400	0.3992	0.0154	0.0139	0.0002	0.930	1.000
Y1	ON							
F&1		0.600	0.5980	0.0216	0.0218	0.0005	0.970	1.000
Y2	ON							
F&1		0.600	0.6020	0.0207	0.0220	0.0004	0.990	1.000
Y3	ON							
F&1		0.600	0.5999	0.0208	0.0219	0.0004	0.980	1.000
Y4	ON							
F&1		0.600	0.6006	0.0215	0.0219	0.0005	0.940	1.000
Y5	ON							
F&1		0.600	0.5985	0.0231	0.0219	0.0005	0.920	1.000

Note that this is counterintuitive from SEM perspective, but not from time series perspective. The model is essentially a factor analysis model with ARMA(1,1) factor

DAFS-WNFS Combo Lag 1 model - ARMA(1,1) factor



Summary of literature: my favorite (read in this order)

- Time series, ARMA literature, econometrics textbooks. W.H. Green, ”Econometric Analysis”
- Comparisons of Four Methods for Estimating a Dynamic Factor Model (2008), by Zhang Hamaker Nesselroade
- Bayesian Estimation of Categorical Dynamic Factor Models (2007), by Zhang and Nesselroade
- On the Nature of SEM Estimates of ARMA Parameters (2002) by Hamaker, Dolan, Molenaar
- A Multilevel AR(1) Model: Allowing for Inter-Individual Differences in Trait-Scores, Inertia, and Innovation Variance (2015) by Jongerling, Laurenceau, Hamaker
- Conditions for the Equivalence of the Autoregressive Latent Trajectory Model and a Latent Growth Curve Model With Autoregressive Disturbances (2005) by Hamaker
- A comparison of Inverse-Wishart prior specifications for covariance matrices in multilevel autoregressive models (2016) by Schuurman, Grasman, Hamaker

Subject-specific times of observations

- The basic model assumes that observations are taken at equally spaced time.
- If times are subject-specific we slice the time grid in sufficiently refined grid and enter missing data for the times where observation is not taken.
- For example if several observations are taken during the day, and at different times for each individual, we slice the day in 24 hour periods and place the corresponding observations in the hour slots.
- Data from the next simulation looks like this for day 1 for individual 1 and 2.

Subject-specific times of observations: subject 1 day 1

y1	y2	y3	y4	y5	T	ID
999	999	999	999	999	1	1
999	999	999	999	999	2	1
999	999	999	999	999	3	1
999	999	999	999	999	4	1
999	999	999	999	999	5	1
999	999	999	999	999	6	1
999	999	999	999	999	7	1
999	999	999	999	999	8	1
999	999	999	999	999	9	1
999	999	999	999	999	10	1
999	999	999	999	999	11	1
999	999	999	999	999	12	1
5.026193	0.327383	1.017519	0.701296	-0.55917	13	1
999	999	999	999	999	14	1
1.628885	1.652829	2.324074	1.800932	4.013447	15	1
999	999	999	999	999	16	1
4.376545	1.652831	2.098822	6.188234	2.913506	17	1
1.534865	0.631455	-0.29779	2.798775	1.37025	18	1
0.359654	1.476764	-0.43374	0.348777	1.382437	19	1
999	999	999	999	999	20	1
999	999	999	999	999	21	1
999	999	999	999	999	22	1
999	999	999	999	999	23	1
999	999	999	999	999	24	1

Subject-specific times of observations: subject 2 day 1

y1	y2	y3	y4	y5	T	ID
999	999	999	999	999	1	2
-4.10343	-2.63665	-0.71492	-1.18325	-2.38467	2	2
999	999	999	999	999	3	2
-1.49744	0.268575	-1.94601	-0.70893	-2.5558	4	2
999	999	999	999	999	5	2
999	999	999	999	999	6	2
999	999	999	999	999	7	2
999	999	999	999	999	8	2
0.158731	-0.49435	2.052861	2.387179	0.932945	9	2
999	999	999	999	999	10	2
999	999	999	999	999	11	2
999	999	999	999	999	12	2
999	999	999	999	999	13	2
999	999	999	999	999	14	2
999	999	999	999	999	15	2
999	999	999	999	999	16	2
999	999	999	999	999	17	2
999	999	999	999	999	18	2
-0.2193	-2.14029	0.050116	-0.42015	-1.00061	19	2
999	999	999	999	999	20	2
999	999	999	999	999	21	2
999	999	999	999	999	22	2
999	999	999	999	999	23	2
999	999	999	999	999	24	2

Subject-specific times of observations - simulation study

```
montecarlo:
    names = y1-y5 u;
    NOBS = 30000;
    NREP = 100;
    NCSIZES = 1;
    CSIZES = 100(300);
    categorical=u;
    generate=u(1);
    within=u;
    missing=y1-y5;

model missing: [y1-y5@-15]; y1-y5 on u@30;

ANALYSIS: TYPE IS TWOLEVEL; estimator=bayes;
          biter=10000(500); proc=2;

model montecarlo:
    %within%
    [u$1*-0.83];
    f by y1-y5*1 (& 1);
    y1-y5*1;
    f@1;
    f on f&1*0.4;

    %between%
    fb by y1-y5*0.5; fb@1; y1-y5*0.2;
```

Subject-specific times of observations - simulation study results

- 80% missing values, 20% present, in hourly scale that means 4 or 5 observations a day
- 99% convergence rate, 96 minutes computation for 100 replications, 1 min per replication

		ESTIMATES		S. E.	M. S. E.	95%	% Sig
	Population	Average	Std. Dev.	Average		Cover	Coeff
Within Level							
F BY							
Y1	1.000	1.0015	0.0166	0.0180	0.0003	0.990	1.000
Y2	1.000	1.0017	0.0172	0.0178	0.0003	0.949	1.000
Y3	1.000	1.0001	0.0183	0.0179	0.0003	0.960	1.000
Y4	1.000	1.0034	0.0157	0.0180	0.0003	0.970	1.000
Y5	1.000	1.0010	0.0192	0.0180	0.0004	0.939	1.000
F ON							
F&1	0.400	0.3981	0.0234	0.0211	0.0005	0.889	1.000

Subject-specific times of observations - summary of results

- Similar results for 85% missing and 90% missing, however, convergence is visibly slower for 90% missing values than for 80% missing values and non-convergence is more likely
- It appears that optimal time discretization would be about 80% to 85% missing values inserted to represent the distance between the observations
- Very likely any information contained in the unequal distances in the observations would be extracted well using the 80% to 85% missing values.
- It appears that this problem is solved with this simple setup and there is no need to develop Brownian motion continuous modeling.
- From a practical point of view it is probably better to develop utilities to help setup the missing data based on the individual-specific and unequally spaced times of observations
- listwise = keepall; new Mplus command to keep the observations with all missing (not needed in the simulation study due to U)

Ignoring subject-specific times of observations

- Using one generated data set from the 80% missing simulation. Consider the results for the autocorraltion parameter under two analysis. True value=0.4. Note that AR meaning depends on the time interval between observations.
- Using listwise = keepall; accounting for subject-specific times.

		Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%
F	ON					
F&1		0.412	0.021	0.000	0.374	0.447

- Without listwise = keepall; ignoring the unequally spaced observations and no missing data: AR is biased

		Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5
F	ON					
F&1		0.103	0.015	0.000	0.071	0.133

Why do we need lag modeling for the latent variables?

- What happens if we just model the lag of the observed sum score? Generate large data set 500 individuals with 100 time points and 3 indicators

```
montecarlo:
    names = y1-y3;
    NOBS = 50000;
    NREP = 1;
    NCSIZES = 1;
    CSIZES = 500(100);
    save=OL.dat;

ANALYSIS:   TYPE IS TWOLEVEL;
            estimator=bayes;
            biter=(500); proc=2;

model montecarlo:
    %within%
    f by y1-y3*1 (& 1);
    y1-y3*1;
    f@1;
    f on f&1*0.5;

    %between%
    fb by y1-y3*0.5; fb@1; y1-y3*0.2;
```

Two-level AR(1)

- No latent variables. Y_{it} is the observed value for individual i at time t.

$$Y_{it} = \mu + Y_i + F_{it}$$

$$F_{it} = \rho F_{i,t-1} + \varepsilon_{it}$$

- The model has just 4 parameters: two variances, one autocorrelation parameter, and one mean parameter.

Two-level AR(1) for the sum score

```
variable:
    names = y1-y3 C;
    usevar= y;
    cluster=c;

define: y=y1+y2+y3;

data: file=OL.dat;

ANALYSIS:   TYPE IS TWOLEVEL;
            estimator=bayes; thin=10;
            biter=(500); proc=2;

model:
    %within%
    f by y@1 (& 1); y@0.01;
    f on f&1*0.5;

    %between%
    y;
```

- The Mplus estimation does require Y@0.01; specification instead of Y@0;. It is an approximation. The closer to zero Y@ is the better the approximation is but the slower the convergence. Recommended value is 1% of the variance of Y.

Why do we need lag modeling for the latent variables?

- Two-level DAFS (true value = 0.5)

		Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%
F	ON					
F&1		0.502	0.005	0.000	0.492	0.510

- Two-level AR(1) for the sum score (bias due to not accounting for measurement error)

		Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%
F	ON					
F&1		0.401	0.004	0.000	0.392	0.409

Three-level AR(1) model (three level univariate model, no latent variables)

- Y_{idt} is the observed value for individual i on day d at time t

$$Y_{idt} = \mu + Y_i + E_{it} + F_{id} + G_{idt}$$

$$G_{idt} = \rho_1 G_{id,t-1} + \varepsilon_{1,idt}$$

$$F_{id} = \rho_2 F_{i,d-1} + \varepsilon_{2,id}$$

- Two type of autocorrelation parameter, ρ_1 is the autocorrelation within the day, ρ_2 is the autocorrelation between the days
- Maybe take out E_{it}?
- Model has 7 parameters: 4 variances, 2 autocorrelations, 1 intercept
- Data consists of 100 individuals, observed for 100 days, with 10 observations per day

Three-level AR(1) model - simulation study

```
montecarlo:
    names = y1-y10;
    NOBS = 10000;
    NREP = 100;
    NCSIZES = 1;
    CSIZES = 100(100);

ANALYSIS:   TYPE IS TWOLEVEL;
estimator=bayes; biter=(500); proc=2;

model montecarlo:
    %within%
    f by y1-y10@1 (& 1);
    y1-y10@0.01;
    f1 by y1@1; f2 by y2@1; f3 by y3@1;
    f4 by y4@1; f5 by y5@1; f6 by y6@1;
    f7 by y7@1; f8 by y8@1; f9 by y9@1;
    f10 by y10@1;
    f1-f10*1 (1);
    f*0.5;
    f on f&1*0.3;
    f2-f10 pon f1-f9*0.5 (2);
    f with f10@0;

    %between%
    fb by y1-y10@1; fb*0.4; y1-y10*0.1;
    [y1-y10*0] (3);
```

Three-level AR(1) model - simulation study results

	Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Within Level							
F ON							
F&1	0.300	0.2955	0.0133	0.0150	0.0002	0.970	1.000
F2 ON							
F1	0.500	0.4993	0.0042	0.0040	0.0000	0.940	1.000
Variances							
F1	1.000	1.0007	0.0049	0.0049	0.0000	0.930	1.000
Residual Variances							
F	0.500	0.4994	0.0150	0.0126	0.0002	0.880	1.000
Between Level							
Intercepts							
Y1	0.000	0.0037	0.0855	0.0572	0.0072	0.810	0.190
Variances							
FB	0.400	0.4310	0.0570	0.0658	0.0042	0.970	1.000
Residual Variances							
Y1	0.100	0.1007	0.0058	0.0051	0.0000	0.940	1.000

Three-level AR(1) model with subject-specific times of observations

- Using 50% missing data. Approximately 5 randomly spaced times of observations per day
- 5 observations a bit too low to obtain good autocorrelation parameter. Sequence is too short?
- Add the commands:
  missing=y1-y10;
  model missing: [y1-y10*0];

Three-level AR(1) model with subject-specific times of observations - simulation results

	Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Within Level							
F ON							
F&1	0.300	0.2864	0.0210	0.0166	0.0006	0.810	1.000
F2 ON							
F1	0.500	0.4428	0.0540	0.0188	0.0062	0.360	1.000
Variances							
F1	1.000	1.0444	0.0429	0.0159	0.0038	0.450	1.000
Residual Variances							
F	0.500	0.4694	0.0307	0.0189	0.0019	0.560	1.000
Between Level							
Intercepts							
Y1	0.000	0.0097	0.0665	0.0589	0.0045	0.890	0.110
Variances							
FB	0.400	0.3821	0.0640	0.0608	0.0044	0.940	1.000
Residual Variances							
Y1	0.100	0.0925	0.0098	0.0057	0.0002	0.660	1.000

Dynamic Latent Class Analysis

- Similar to DSEM. U_{it} is a vector of categorical variable for individual i at time t measuring a within latent class variable C_{it} and a between level latent class variable D_i
- Two-level LCA (Vermunt formulation, simpler than the general Mplus model)

$$P(U_{it1},...,U_{itJ}|C_{it}) = P(U_{it1}|C_{it})...P(U_{itJ}|C_{it})$$

$$P(C_{i1} = k_1|D_i = l) = E_{k_1,l}, P(D_i = l) = Q_l$$

$$P(C_{it} = k_1|C_{i,t-1} = k_2, D_i = l) = P_{k_1,k_2,l}$$

- $E_{k_1,l}$, $P_{k_1,k_2,l}$ and Q_l are probabilities with Dirichlet prior. Transition matrix is specific for each between level class. Easy MCMC estimation.