

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

Tihomir Asparouhov  
Part 6

August 3, 2017

- DSEM output options and plots
- Subject-specific variances
- Unevenly spaced and individual-specific times of observations
- Two-level DAFS and WNFS
- TVEM - time varying effects models
- Three level AR(1) models: within day v.s. between day autoregressive modeling

# DSEM output options

- residual option: model estimated means, variance and autocorrelations for the observed variables
- residual(cluster) option: model estimated and cluster/subject specific means, variance and autocorrelations for the observed variables
- tech4 and tech4(cluster) options: model estimated quantities for the latent variables
- stand and stand(cluster) options: standardized model estimates and standardized cluster specific model estimates
- The option with (cluster) also provides the average across cluster quantities for the cluster specific estimates - applies for residual/tech4/stand
- The (cluster) option new also for none-DSEM models
- All of the above are based on Yule-Walker and require stationarity of the autoregressive part of the model
- HTML clickable output

# DSEM output example

---

|           |   |
|-----------|---|
| VARIABLE: | NAMES = y x1 x2 c;<br>WITHIN = x1;<br>BETWEEN = x2;<br>CLUSTER = c;<br>LAGGED = y(1); |
| DATA:     | FILE = a.dat;   |
| ANALYSIS: | TYPE = TWOLEVEL RANDOM;<br>ESTIMATOR = BAYES;<br>PROCESSORS = 2;                      |
| MODEL:    | %WITHIN%<br>s1   y ON x1;<br>s2   y ON y&1;<br>s3   y;<br>%BETWEEN%<br>y s1-s3 ON x2; |
| OUTPUT:   | STANDARDIZED(CLUSTER) RESIDUAL(CLUSTER) FSCOMPARISON;                                 |
| PLOT:     | TYPE = PLOT3;   |
| SAVEDATA: | STDDISTRIBUTION = 1.dat;<br>SAVE = fs(200);<br>FILE = 2.dat;<br>BPARAMETER = 3.dat;   |

---

# DSEM output example: htm output



Mplus - [gb00o.htm]

File Edit View Mplus Plot Diagram Window Help



Font size: A A A A

Mplus VERSION 8  
MUTHEN & MUTHEN  
07/05/2017 10:58 AM

## OUTPUT SECTIONS

Input Instructions  
Input Warnings And Errors  
Summary Of Analysis  
Summary Of Data  
Covariance Coverage Of Data  
Univariate Sample Statistics  
Model Warnings And Errors  
Model Fit Information  
Model Results  
Standardized Model Results  
R-square  
Within-level Standardized Model Results For Cluster 1  
Within-level R-square For Cluster 1  
Within-level Standardized Model Results For Cluster 2  
Within-level R-square For Cluster 2  
Within-level Standardized Model Results For Cluster 3

# DSEM output example: standardized model results

## STDYX Standardization

|  | Estimate | Posterior<br>S.D. | One-Tailed<br>P-Value | 95% C.I.   |            | Significance |
|--|----------|-------------------|-----------------------|------------|------------|--------------|
|  |          |                   |                       | Lower 2.5% | Upper 2.5% |              |
| Within-Level Standardized Estimates Averaged Over Clusters |          |                   |                       |            |            |              |
| S1   Y ON  |          |                   |                       |            |            |              |
| X1   | 0.438    | 0.007             | 0.000                 | 0.427      | 0.454      | *            |
| S2   Y ON  |          |                   |                       |            |            |              |
| Y&1  | 0.247    | 0.007             | 0.000                 | 0.232      | 0.260      | *            |
| S3   |          |                   |                       |            |            |              |
| Y  | 0.452    | 0.009             | 0.000                 | 0.427      | 0.464      | *            |
| Between Level  |          |                   |                       |            |            |              |
| S1 ON  |          |                   |                       |            |            |              |
| X2   | 0.359    | 0.065             | 0.000                 | 0.208      | 0.459      | *            |
| S2 ON  |          |                   |                       |            |            |              |
| X2   | 0.682    | 0.045             | 0.000                 | 0.586      | 0.758      | *            |
| S3 ON  |          |                   |                       |            |            |              |
| X2   | 0.335    | 0.059             | 0.000                 | 0.192      | 0.442      | *            |
| Means  |          |                   |                       |            |            |              |
| Y  | 0.181    | 0.103             | 0.035                 | -0.005     | 0.393      |              |
| Intercepts   |          |                   |                       |            |            |              |
| S1   | 0.898    | 0.118             | 0.000                 | 0.676      | 1.111      | *            |
| S2   | 1.219    | 0.114             | 0.000                 | 0.969      | 1.406      | *            |

# DSEM output example: cluster specific standardized results

Mplus - [gb00o.htm]

File Edit View Mplus Plot Diagram Window Help



## WITHIN-LEVEL R-SQUARE FOR CLUSTER 1

| Variable | Estimate | Posterior | One-Tailed | 95% C.I.   |            |
|----------|----------|-----------|------------|------------|------------|
|          |          | S.D.      | P-Value    | Lower 2.5% | Upper 2.5% |
| Y        | 0.607    | 0.097     | 0.000      | 0.460      | 0.871      |

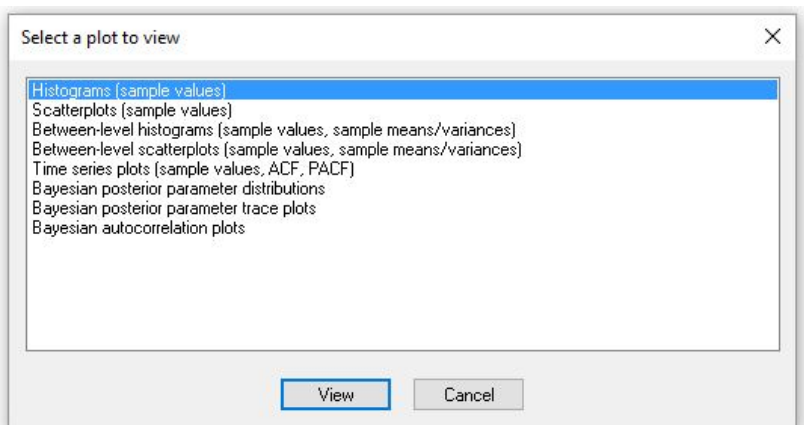
## WITHIN-LEVEL STANDARDIZED MODEL RESULTS FOR CLUSTER 2

### STDYX Standardization

|                  | Estimate | Posterior | One-Tailed | 95% C.I.   |            | Significance |
|------------------|----------|-----------|------------|------------|------------|--------------|
|                  |          | S.D.      | P-Value    | Lower 2.5% | Upper 2.5% |              |
| S1   Y ON<br>X1  | 0.539    | 0.064     | 0.000      | 0.398      | 0.647      | *            |
| S2   Y ON<br>Y&1 | 0.250    | 0.070     | 0.000      | 0.108      | 0.388      | *            |
| S3  <br>Y        | 0.610    | 0.075     | 0.000      | 0.471      | 0.768      | *            |

STDY Standardization

# DSEM plots: plot menu





# DSEM plots: plotting model estimated v.s. observed cluster specific statistic

Between-level scatter plots

Plot properties

Variables selection: (see notations below)

X: Y (variance over Within) Y: Y (estimated cluster variance)

View properties:

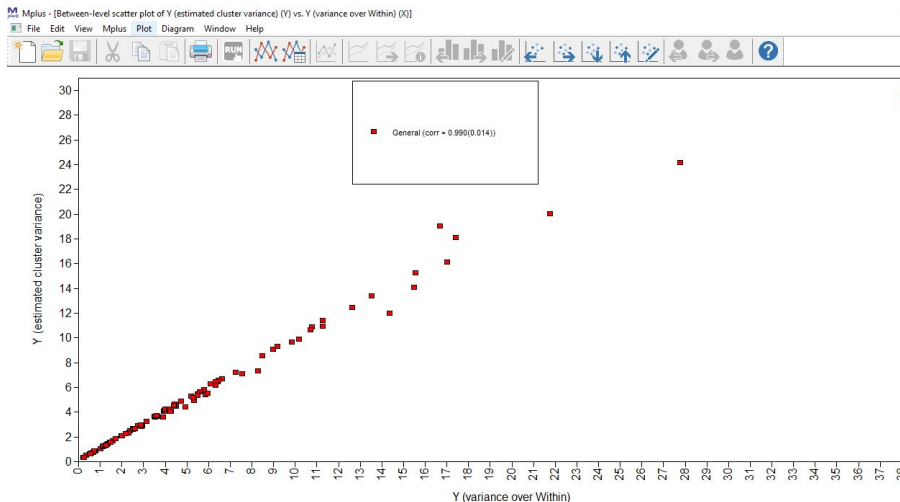
- ☒ Show all values (groups/classes separated by color)
- ☐ Show all values (groups/classes not separated)
- ☐ Show only specific group/class
- ☐ Show values by cluster

Group/class selection:

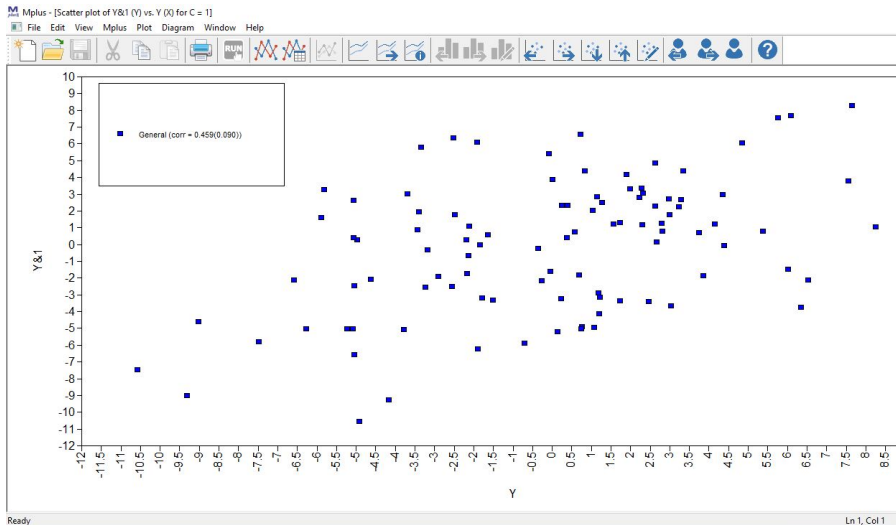
General

Notations: For estimated factor scores, "mean" and "median" denote the mean and median of the latent variable distribution.

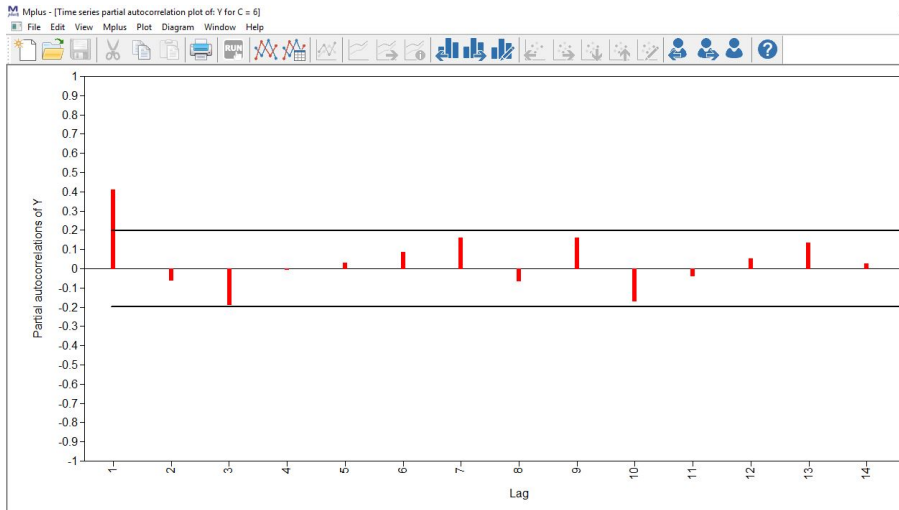
# DSEM plots: plotting model estimated v.s. observed cluster specific variances



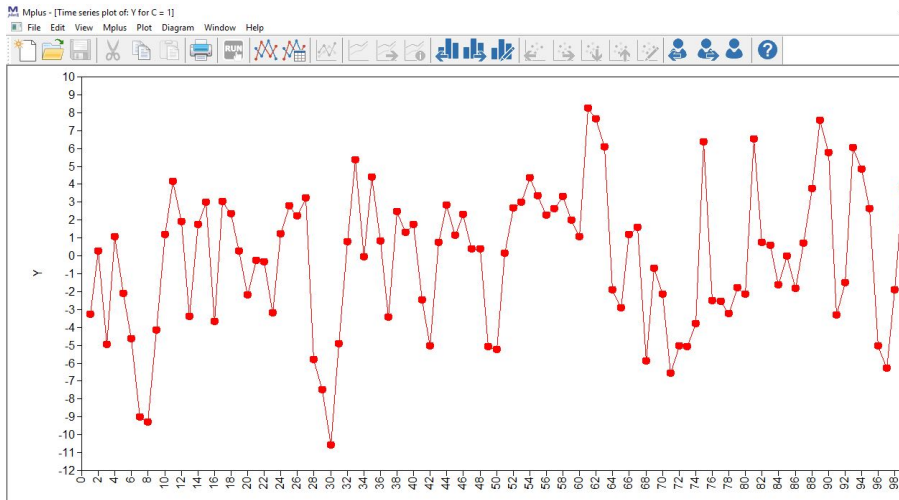
# DSEM plots: cluster specific plots



# DSEM plots: subject specific partial autocorrelation function



# DSEM plots: subject specific time series plots



- DSEM output options and plots
- **Subject-specific variances**
- Unevenly spaced and individual-specific times of observations
- Two-level DAFS and WNFS
- TVEM - time varying effects models
- Three level AR(1) models: within day v.s. between day autoregressive modeling

- Jongerling J, Laurenceau J.P., Hamaker E. (2015). A Multilevel AR(1) Model: Allowing for Inter-Individual Differences in Trait-Scores, Inertia, and Innovation Variance. Multivariate Behav Res. 50(3), 334-349.
- In this paper it is shown that if subject specific variances are ignored - the structural parameters can be slightly biased. This does not happen in regular two-level models.

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

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

$$v_i = \text{Log}(\text{Var}(\xi_{it}))$$

- The bias depends on how high the correlation is between  $r_i$  and  $v_i$

# Subject specific variance -results

**Table:** Comparing the estimation with random variance and without random variance (non-random variance): Bias(coverage)

| parameter  | $Cov(r_i, v_i)$ | random variance | non-random variance |
|------------|-----------------|-----------------|---------------------|
| $E(r_i)$   | high            | .001(.97)       | .040(.35)           |
| $E(r_i)$   | medium          | .001(.98)       | .028(.65)           |
| $E(r_i)$   | low             | .001(.97)       | .017(.83)           |
| $E(r_i)$   | none            | .001(.96)       | .007(.92)           |
| $Var(r_i)$ | high            | .001(.97)       | -.012(.47)          |
| $Var(r_i)$ | medium          | .001(.93)       | -.007(.78)          |
| $Var(r_i)$ | low             | .001(.93)       | -.004(.88)          |
| $Var(r_i)$ | none            | .001(.94)       | -.001(.91)          |



More detailed method for evaluation of model estimation

$$SMSE = \sqrt{(1/N) \sum_i (\hat{r}_i - r_i)^2}$$

| $Cov(r_i, v_i)$ | random variance | non-random variance |
|-----------------|-----------------|---------------------|
| high            | .255            | .346                |
| medium          | .293            | .329                |
| low             | .300            | .316                |
| none            | .300            | .310                |

# Subject specific variance - conclusions

- Looking at the parameter estimates alone may not be enough when comparing estimation methods. Distortion of structural parameters due to ignoring subject specific variance is not simple shift in the autoregressive parameter. Error is actually doubled when looking at the random effects SMSE.
- Even in standard two-level models, using cluster specific variance is important if we use SMSE as a criterion
- Subject specific variance extracts more information from the data, yields more accurate estimation
- More simulations are needed to evaluate this issue in multivariate setting - study the effect of subject specific covariance.

- DSEM output options and plots
- Subject-specific variances
- **Unevenly spaced and individual-specific times of observations**
- Two-level DAFS and WNFS
- TVEM - time varying effects models
- Three level AR(1) models: within day v.s. between day autoregressive modeling

# 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.

# 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 - simulation study

---

|                   |  |
|-------------------|--|
| MONTECARLO:       | NAMES = y1-y5 u;<br>NOBSERVATIONS = 30000;<br>NREPS = 100;<br>NCSIZES = 1;<br>CSIZES = 100(300);<br>CATEGORICAL = u; ! u is used to give the same missingness for all y1-y5<br>GENERATE = u(1);<br>WITHIN = u;<br>MISSING = y1-y5; |
| MODEL MISSING:    | [y1-y5@-15];<br>y1-y5 ON u@30;   |
| ANALYSIS:         | TYPE = TWOLEVEL;<br>ESTIMATOR = BAYES;<br>BITERATIONS = 10000(500);<br>PROCESSORS = 2;   |
| MODEL MONTECARLO: | %WITHIN%<br>[u\$1*-0.83]; ! Probit of -.83 gives 20% present and 80% missing<br>f BY y1-y5*1 (& 1);<br>y1-y5*1;<br>f ON f&1*0.4;<br>%BETWEEN%<br>fb BY y1-y5*0.5;<br>fb@1;<br>y1-y5*0.2;   |

---

**Table:** Two-level DAFS AR(1) with subject-specific times - simulation study results

| percentage missing values | $\hat{\phi}$ (coverage)<br>$\phi = 0.4$ | convergence rate | comp time per replication in min |
|---------------------------|---|------------------|----------------------------------|
| .80                       | .39(.95)                                | 100%             | 1.5                              |
| .85                       | .39(.90)                                | 95%              | 2.5                              |
| .90                       | .35(.46)                                | 55%              | 10                               |
| .95                       | .34(.55)                                | 55%              | 18                               |

Quality of the estimation deteriorates as the amount of inserted missing data exceeds 90%. Note that this missing data is imputed by the MCMC estimation, leading to large amount of imputed quantities. It works well with 80% and 85% missing data.

# Subject-specific times of observations

- Information contained in the unequal distances in the observations would be extracted well using the 80% to 85% missing values, eliminating the need for continuous time modeling
- Tinterval command will setup the missing data for you, given the precise times of observations and an approximation value  $\delta$
- Tinterval =  $t(\delta)$  means that the continuous time variable  $t$  is replaced by the nearest integer  $[t/\delta]$ . There are complications if the nearest integers is the same for two or more different observations times  $t$ . Special algorithm to resolve this issue.

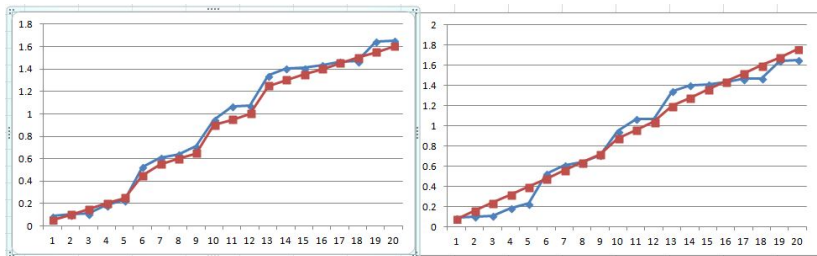


# Subject-specific times of observations

- Split the time axis in bins of size  $\delta$ . Then place each observation in the correct bin. Repeat these steps until each bin contains no more than 1 observation
  - find a bin with more than 1 observations
  - locate the nearest empty bin (look up or down)
  - move one of the extra observation to fill in the the empty bin but keep order of the observations so the extra observation bumps the rest of the observations towards the empty bin
- Mplus will warn you if the shifting process yields a discrepancy between  $t/\delta$  and new time bigger than 5. Lower the  $\delta$  value to resolve this.
- Fill in the remaining bins with missing values and set the time as  $T=1,2, \dots$  and  $T$  is the bin number.
- Other algorithms are possible. Make your own discretization algorithm and use Mplus with integer times.

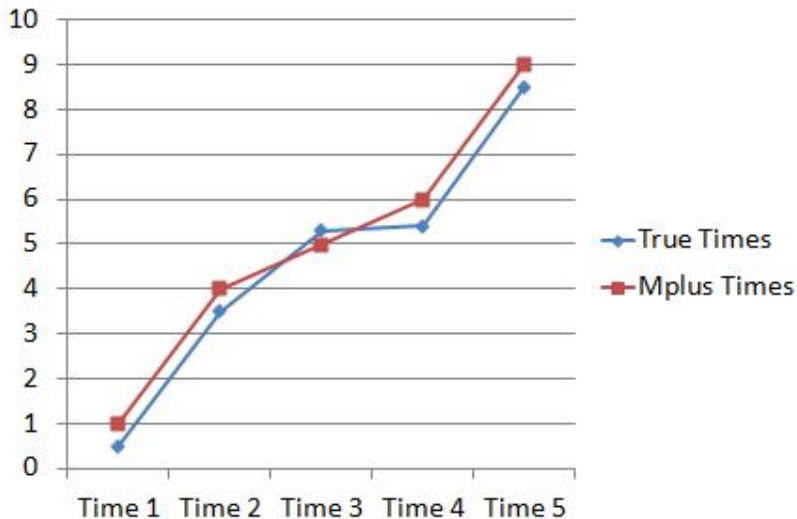
# Tinterval command comparison

Tinterval(0.05) v.s. Tinterval(0.08), Blue=true times, Red=Mplus generated times



Similar plots can be produced with real data sets to evaluate the Tinterval approximation for real data.

# Tinterval command illustration



# Simulation study with varying $\delta$

**Table:** Two-level AR(1) with subject-specific times. Estimates and coverage for  $\phi$  and amount of missing data  $m_2$  during the analysis,  $\text{tinterval}=\delta$

| $m$ | $\delta$ | $\phi = 0.8$ | $m_2$ |
|-----|----------|--------------|-------|
| .80 | 1        | .80(.91)     | .80   |
| .80 | 2        | .81(.31)     | .58   |
| .80 | 3        | .83(.00)     | .38   |
| .80 | 4        | .84(.00)     | .18   |
| .80 | 5        | .86(.00)     | .05   |
| .80 | 10       | .92(.00)     | .00   |
| .95 | 1        | .80(.85)     | .95   |
| .95 | 2        | .81(.57)     | .90   |
| .95 | 3        | .82(.20)     | .85   |
| .95 | 4        | .83(.00)     | .80   |
| .95 | 5        | .84(.00)     | .74   |
| .95 | 10       | .88(.00)     | .49   |

# Subject-specific times of observations

- Tinterval command is not perfect. It is an approximate solution for the continuous process.
- The main question is how to choose  $\delta$ . Three considerations:
  - Choose scale that is natural to help with interpretation of model and results - hour, day, week
  - Choose scale that does not produce more than 90% missing data, around 80% unless lower is appropriate
  - Smaller values yield better approximations but also more missing data
  - TVEM models / Cross-classified DSEM: small  $\delta$  will lead to too many time specific random effects

- DSEM output options and plots
- Subject-specific variances
- Unevenly spaced and individual-specific times of observations
- **Two-level DAFS and WNFS**
- TVEM - time varying effects models
- Three level AR(1) models: within day v.s. between day autoregressive modeling

# Simulation Study - Twolevel 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;
```

# Results - Twolevel 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 |

Note that the choice of these coefficients for simulation study purposes is tricky as you need to preserve the stationarity of the model, otherwise the data will explode. Simple rule: if all are positive and the sum is less than 1 the process is stationary. The exact stationarity condition involves finding the roots of a polynomial of degree  $L$ .



# Simulation Study - Twolevel 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;
```

# Results - Twolevel 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 |

Note that the choice of these coefficients for simulation study purposes is NOT tricky: WNFS is always stationary.

# Simulation Study - Twolevel 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;
```

# Results - Twolevel 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

- DSEM output options and plots
- Subject-specific variances
- Unevenly spaced and individual-specific times of observations
- Two-level DAFS and WNFS
- **TVEM - time varying effects models**
- Three level AR(1) models: within day v.s. between day autoregressive modeling

# Two-level MEAR(1) TVEM - simulated example

- General TVEM framework: multivariate, multilevel, time-series, continuous and categorical dependent variables.
- Consider the following example  $N=500$ ,  $T=50$ : Two-level MEAR(1) with covariate with time specific mean and regression coefficient

$$Y_{it} = \mu_t + Y_i + \beta_t X_{it} + f_{it} + \varepsilon_{it}$$

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

$$\mu_t = f_1(t) = \log(t)$$

$$\beta_t = f_2(t) = a + bt + ct^2 = 0.001 \cdot t \cdot (50 - t)$$

- $f_1(t)$  and  $f_2(t)$  are arbitrary functions of  $t$  used for data generation

- Non-parametric cross-classified: Exploratory TVEM-DSEM, no parametric shape included in the model. The random effect  $\mu_t$  and  $\beta_t$  are modeled as normally distributed random effects
- Semi-parametric cross-classified: TVEM-DSEM, parametric shape included in the model as well as residual random effects. The random effects  $\mu_t$  and  $\beta_t$  are modeled as normally distributed random effects which include parametric curves for the means of the random effects
- Parametric two-level: TVEM-DSEM, parametric shape included in the model. The random effects  $\mu_t$  and  $\beta_t$  are modeled as normally distributed random effects with zero variance, only parametric curve for the the random effect mean. Typically such a model is estimated as two-level DSEM rather than cross-classified.

```
variable:
names are y x t c;
within=x;
cluster=c t;

analysis: type = cross random;
estimator=bayes; biter=(5000); process=2;

model:
%within%
y*0.5; eta by y@1 (&1);
eta on eta&1*0.5; eta*1.2;
s | y on x;

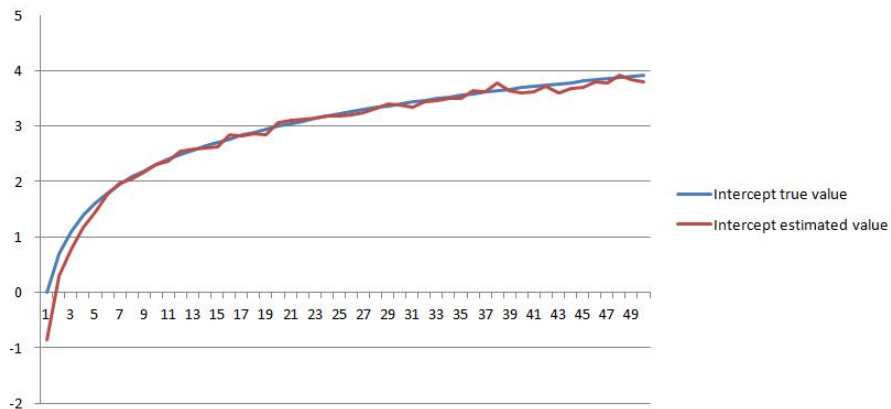
%between t%
y; s; [s*0];

%between c%
y*0.5; [y*0]; s@0; [s@0];

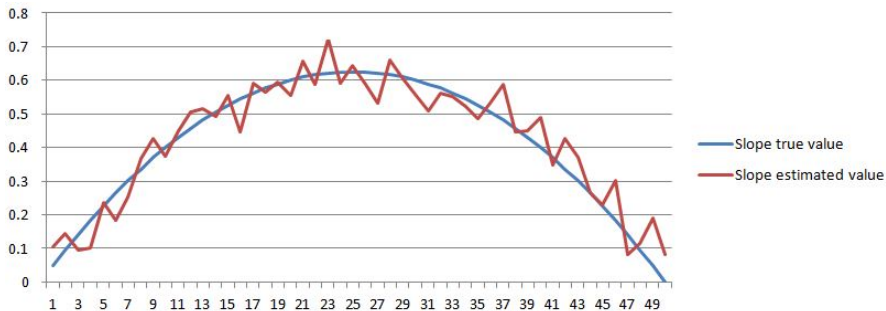
savedata: file is TVEM3.dat; save=FS(2000);
```



# Exploratory TVEM-DSEM results



# Exploratory TVEM-DSEM results



# Semi-parametric TVEM-DSEM

```
variable:
names are y x t c;
within=x;
cluster=c t;
usevar=y x x1 x2 x3;
between=(t) x1 x2 x3;

define: x1=log(t); x2=0.05*t; x3=0.001*t*t;

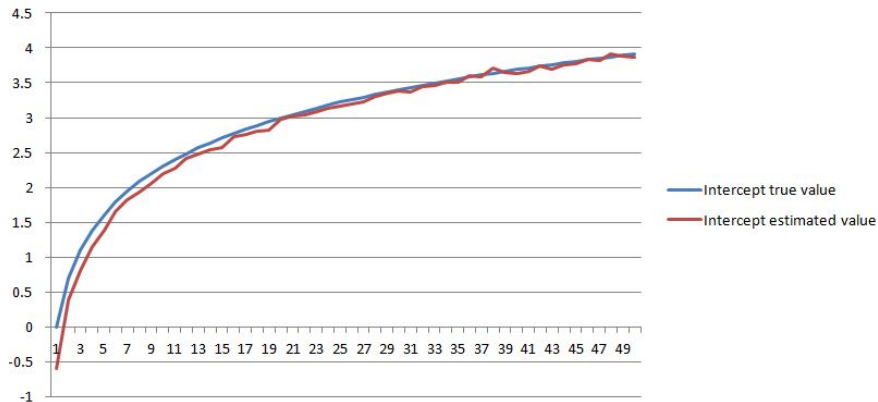
analysis: type=cross random;
estimator=bayes; biter=(5000); process=2;

model:
%within%
y*0.5; eta by y@1 (&1);
eta on eta&1*0.5; eta*1.2;
s | y on x;

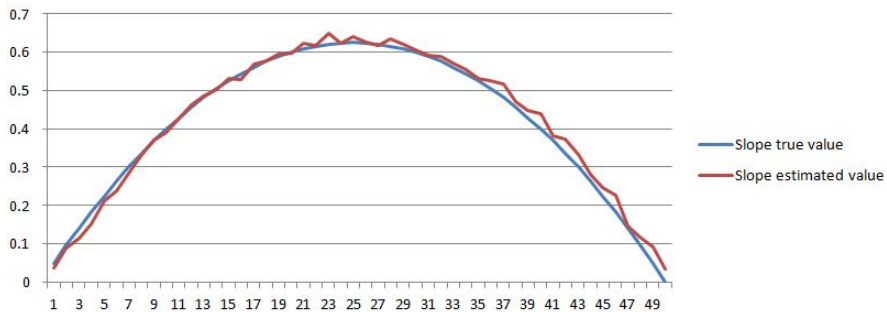
%between t%
y; s; [s*0]; y on x1*1; s on x2*1 x3*-1;

%between c%
y*0.5; [y*0]; s@0; [s@0];
```

# Semi-parametric TVEM-DSEM results



# Semi-parametric TVEM-DSEM results



# Parametric: TVEM-DSEM - two level model

```
variable:
names are y x t c;
within=x x1 x2 x3;
cluster=c;
usevar=y x x1 x2 x3;

define: x1=log(t); x2=0.05*t*x; x3=0.001*t*t*x;

analysis: type = twolevel random;
estimator=bayes; biter=(10000); process=2;

model:
%within%
y*0.5; eta by y@1 (&1);
eta on eta&1*0.5; eta*1.2;
y on x*0 x1*1 x2*1 x3*-1;

%between%
y*0.5; [y*0];
```

- DSEM output options and plots
- Subject-specific variances
- Unevenly spaced and individual-specific times of observations
- Two-level DAFS and WNFS
- TVEM - time varying effects models
- **Three level AR(1) models: within day v.s. between day autoregressive modeling**

# Three-level AR(1) model

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

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

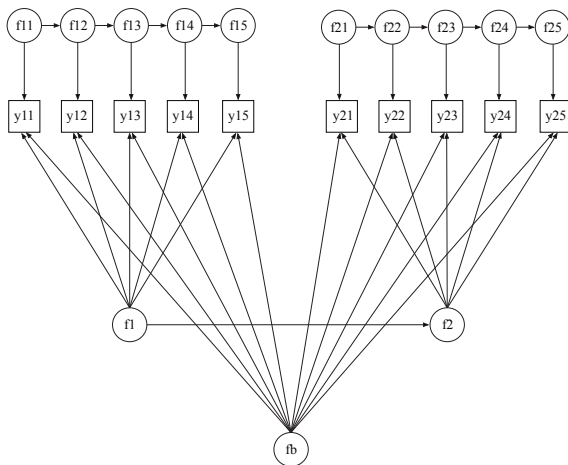
$$G_{idt} = \rho_1 G_{id,t-1} + \varepsilon_{1,idt}$$

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

- Two type of autocorrelation parameter,  $\rho_1$  is the autocorrelation within the day,  $\rho_2$  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
- Model the additional level in wide format either on within or between (whichever has smaller width)



# Three-level AR(1) model



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

---

|                   |   |
|-------------------|---|
| MONTECARLO:       | NAMES = y1-y10;<br>NOBSERVATIONS = 10000;<br>NREPS = 100;<br>NCSIZES = 1;<br>CSIZES = 100(100);   |
| ANALYSIS:         | TYPE = TWOLEVEL;<br>ESTIMATOR=BAYES;<br>BITERATIONS=(500);<br>PROCESSORS=2;   |
| MODEL MONTECARLO: | %WITHIN%<br>f BY y1-y10@1 (&1);<br>y1-y10@0.01;<br>f1 BY y1@1; f2 BY y2@1; f3 BY y3@1;<br>f4 BY y4@1; f5 BY y5@1; f6 BY y6@1;<br>f7 BY y7@1; f8 BY y8@1; f9 BY y9@1;<br>f10 BY y10@1;<br>f1-f1*1 (1);<br>f*0.5;<br>f ON f&1*0.3;<br>f2-f10 pon f1-f9*0.5 (2);<br>f WITH f10@0;<br>%BETWEEN%<br>fb BY y1-y10@1;<br>fb*0.4;<br>y1-y10*0.1;<br>[V1-V10*0] (3); |

---

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

|                    | Population | Estimates<br>Average | Std. Dev. | S.E.<br>Average | M.S.E. | 95%<br>Cover | % Sig<br>Cover |
|--------------------|------------|----------------------|-----------|-----------------|--------|--------------|----------------|
| 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. Needs much longer MCMC estimation.
- 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<br>Average | Std. Dev. | S.E.<br>Average | M.S.E. | 95%<br>Cover | % Sig<br>Cover |
|--------------------|------------|----------------------|-----------|-----------------|--------|--------------|----------------|
| Within Level       |            |                      |           |                 |        |              |                |
| F ON               |            |                      |           |                 |        |              |                |
| F&l                | 0.300      | 0.2864               | 0.0210    | 0.0166          | 0.0006 | 0.810        | 1.000          |
| F2 ON              |            |                      |           |                 |        |              |                |
| Fl                 | 0.500      | 0.4428               | 0.0540    | 0.0188          | 0.0062 | 0.360        | 1.000          |
| Variances          |            |                      |           |                 |        |              |                |
| Fl                 | 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          |

# New models coming in future Mplus release

- Residual DSEM
- Bayesian Multilevel Mixture Models
- Bayesian Multilevel Latent Transition Models with cluster specific transition probabilities
- Multilevel Mixtures of Dynamic Structural Equation Models
- Dynamic Latent Class Analysis
- Multilevel Hidden Markov Model
- Multilevel Markov Switching Autoregressive Models
- Multilevel Markov Switching DSEM Models
- These are at various stages of development.
- Asparouhov, T., Hamaker, E.L. & Muthén, B. (2017). Dynamic Latent Class Analysis, Structural Equation Modeling: A Multidisciplinary Journal, 24:2, 257-269