Three-Level Dynamic Structural Equation Modeling

Tihomir Asparouhov and Bengt Muthén *

October 14, 2025

*We thank Ellen Hamaker for valuable comments on the earlier draft

Abstract

In this paper we describe a three level dynamic structural modeling framework as a generalization of the DSEM framework of Asparouhov et al. (2018). Model estimation is discussed and the framework is illustrated with simulation studies and practical examples. Two common scenarios are described. The first is intensive longitudinal data for a group of individuals where observations are nested within days, periods, waves or bursts. The second is intensive longitudinal data for individuals nested within higher level clusters. Comparisons are made with two-level DSEM models and emphasis is given to what can be learned from the additional level of clustering.

1 Introduction

In this paper, we explore the possibility of expanding the Mplus dynamic structural equation modeling (DSEM) framework of Asparouhov et al. (2018) to accommodate an additional, third, level of nesting. Such a modeling framework can also be viewed as an extension of the standard hierarchical three-level framework for the special case where the lowest level of clustering arises from repeated measures across time. As in DSEM, modeling repeated observations across time for multiple individuals needs to accommodate autocorrelation of the observations across time in addition to the correlation due to nested observations modeled by the random effects in the multilevel models.

DSEM can be viewed as a computationally practical alternative to RI-CLPM (random intercept cross-lagged panel modeling) when the number of repeated observations for an individual is large, such as with intensive longitudinal data (ILD). In that regard, DSEM3 (three-level DSEM) can be viewed as a practical alternative to a two-level RI-CLPM for ILD. Two-level RI-CLPM can be estimated with maximum likelihood estimation in Mplus.

There are several potential application areas for DSEM3. First, when repeated measures are observed for individuals who are also nested within higher-level units such as schools or hospitals, we can use the additional level of nesting to account for school-level or hospital-level effects, while still providing DSEM-level modeling for individual ILD. In such an application, the additional clustering level is the highest level of clustering. DSEM in its nature is profoundly a single-level model despite the fact that we do portray it as a two-level model. That is because we model data for individuals. The two-level nature of the model is just to capture the ILD. Multilevel modeling is essential, however, to study group-level effects on individuals, such as hospital effects, treatment effects, or classroom or school effects. Thus, DSEM3 is essentially multilevel-DSEM, which allows us to study these group-level effects for individuals that have ILD.

The second application area is the case of bursts of ILD for multiple individuals. Study designs may include multiple series of repeated measures. For example, once a month, intense daily data might be collected to study an individual's daily dynamics. Collecting the data over many months may allow us to study the long-term evolution of variables, while the intense daily data would still allow us to study momentary assessment and dynamics. In this design, the new level is the middle level, which is responsible for capturing

the monthly evolution of the variables that are not time-invariant. Subject-specific effects will remain at the highest level.

The third application area is similar to the "burst of ILD" application, however, here the bursts are not based on the design but on the human nature of sleep. Sleep is a fundamental disruptor of momentary assessment models. Ignoring the night-time sleep's ability to reset and restore human behavior in a fundamental way is a detriment. Treating "sleep time" as just another 8-hour period is likely to distort the dynamics we want to study. Continuous time analysis in Asparouhov and Muthén (2014) shows that for one example, sleep-time is equivalent to 1.8 times daytime as far as memory of emotions in concerned. Clearly, however, different emotions would be affected differently by the night-time reset, which thus impairs our ability to study multiple variables. We cannot fix this problem by stretching the time during the night. There is an additional computational issue that DSEM suffers from when accounting for night time. The TINTERVAL command inserts missing data when there are no observations collected, i.e., in an hourly design, 8 rows of missing data are inserted for every day of observation to account for sleep time. That level of missing data can be difficult for the Bayesian estimation, where every missing value is imputed at every iteration. Depending on the complexity of the model, there is an upper limit of missing data that an estimation can handle without being either too slow or too close to convergence problems or identification issues. DSEM3 offers a valuable alternative and potential solution to modeling night time, not just in terms of the computational aspect but also in terms of modeling the fundamental reset. For DSEM3, the middle level will be the day effects, while the lower level will contain the within-day data. The highest level is the individual level. This within-day, between-day, between-person model may in fact be the most common application area for DSEM3 because many existing ILD designs appear to fit this general description.

The fourth application area is in study design. Consider the fundamental question of how often we need to measure a variable. Clearly, most variables do not need to be measured every second, and most momentary assessment studies would need more than one observation a day. DSEM3 can be used to determine the proper study design by analyzing "excessively" collected data for a small group of individuals. Consider as an example the situation where we want to determine if observations should be collected every 5 minutes or every half an hour. The question is if a significant and meaningful change occurs over a 5-minute period that will allow us to study variable dynamics,

or if a 30-minute period is sufficient. To answer this question, we can collect data on a small set of subjects 5 minutes apart. We can then conduct a DSEM3 analysis where the highest level is still the individual, the 30-minute period is the middle level, while the 6 observations within the thirty-minute period consist of the small cluster at the lowest level. The ultimate goal is to decompose the variance of the variables using only univariate analysis. The variance of the variable is decomposed in 3 ways: one for each level. The variance on the lowest level is further decomposed as random shock and autocorrelation. If the total within-level variance is smaller than 1 or 2 percent of the total variance, we can surely conclude that the 30-minute level is sufficient. We can repeat the above analysis for different size periods until we find an interval where the within-level variance is at sufficiently substantial levels. Using DSEM3 for the above procedure is important because if we use a standard 3-level model, the part of the variance due to autocorrelation could be converted to a higher level variance and the within level variance would be underestimated.

The above applications are of two types. The first type is when the individual is the middle level, and individuals are nested within clusters (schools or hospitals). ILD is collected for the individuals. The second type is when the individual is the highest level cluster, while the middle clustering is based on time (such as in the bursts). The two types of applications are quite distinct in a way because when the middle level represents another time variable, there might be a possibility to further model autocorrelation on the middle level. Furthermore, when the middle level represents another time variable, DSEM3 might reduce the autocorrelations as compared to DSEM2. For example, introducing a day-specific effect in a model will compete with the autocorrelation of the within-day observations. The level of reduction in the autocorrelations would affect the cross-lag modeling and might have substantial practical implications.

In what follows, we will continue to motivate DSEM3 with simulation studies and real-data examples. We also discuss a formal definition of the model as well as the model estimation. We limit ourselves to the case of normally distributed variables as well as the DSEM framework alone and not the RDSEM framework which separates predictors and dynamic variables.

2 The DSEM3 model

Let Y_{ijt} be a vector of measurements for individual i in cluster j at time t, where the i-th individual is observed at times $t = 1, 2, ..., T_{ij}$. The DSEM3 model of lag L begins with the following decomposition

$$Y_{ijt} = Y_{1,ijt} + Y_{2,ij} + Y_{3,j}, \tag{1}$$

where $Y_{2,ij}$ and $Y_{3,j}$ are individual-specific and cluster-specific contributions and $Y_{1,tij}$ is the time t specific deviation for individual i in cluster j, i.e., the residual. All three components are latent normally distributed random vectors and are used to form three sets of structural equations - one on each level.

The within-level part of the DSEM3 model is described by the following equations which include time-series components of lag L and is essentially the same as the within-level DSEM model

$$Y_{1,ijt} = \nu_1 + \sum_{k=0}^{L} \Lambda_{1,k}\eta_{1,ij,t-k} + \sum_{k=0}^{L} R_l Y_{1,ij,t-k} + \sum_{k=0}^{L} K_{1,l} X_{1,ij,t-k} + \varepsilon_{1,ijt} \tag{2}$$

$$\eta_{1,ijt} = \alpha_1 + \sum_{k=0}^{L} B_{1,k}\eta_{1,ij,t-k} + \sum_{k=0}^{L} Q_k Y_{1,ij,t-k} + \sum_{k=0}^{L} \Gamma_{1,k} X_{1,ij,t-k} + \xi_{1,ijt}. \tag{3}$$

Here $X_{1,ijt}$ is a vector of observed covariates for individual i in cluster j at time t and $\eta_{1,ijt}$ is a vector of latent variables/factors. The time-series nature of the model is contained in the fact that all of the variables from the previous L periods can be used as predictors as well for the current period. When the time index $t - k$ becomes non-positive, the variable is treated as missing.

The second and third level models are standard structural equation models and are essentially as in a standard 3-level SEM

$$Y_{2,ij} = \nu_2 + \Lambda_2\eta_{2,ij} + K_2 X_{2,ij} + \varepsilon_{2,ij} \tag{4}$$

$$\eta_{2,ij} = \alpha_2 + B_2\eta_{2,ij} + \Gamma_2 X_{2,ij} + \xi_{2,ij} \tag{5}$$

$$Y_{3,j} = \nu_3 + \Lambda_3\eta_{3,j} + K_3 X_{3,j} + \varepsilon_{3,j} \tag{6}$$

$$\eta_{3,j} = \alpha_3 + B_3\eta_{3,j} + \Gamma_3 X_{3,j} + \xi_{3,j}. \tag{7}$$

The vector $X_{2,ij}$ is a vector of individual-specific time-invariant covariates and $X_{3,j}$ is a vector of cluster-specific covariates. Similarly, $\eta_{2,ij}$ is a vector of individual-specific time-invariant latent variables and $\eta_{3,j}$ is a vector of cluster-specific latent variables. The variables $\varepsilon_{1,ijt}$, $\xi_{1,ijt}$, $\varepsilon_{2,ij}, \xi_{2,ij}, \varepsilon_{3,j}, \xi_{3,j}$ are zero mean residuals as usual. Regressions among Y components on the between levels are indirectly included as well by creating a latent variable equal to the Y variable on the corresponding level.

Note that all covariates are not decomposed as the dependent variables are in (1). The covariates in the above model are present only on one of the levels. If a covariate needs to be decomposed, it must be treated as a dependent variable or alternatively it can be decomposed using observed centering. That is, a covariate X_{ijt} can be replaced by three observed centered covariates, one on each level: $\overline{\overline{X_{*j*}}}$, $\overline{X_{ij*}} - \overline{\overline{X_{*j*}}}$ and $X_{ijt} - \overline{X_{ij*}}$. Here $\overline{X_{ij*}}$ is the average over t of X_{ijt} and $\overline{\overline{X_{*j*}}}$ is the average over i of $\overline{X_{ij*}}$. Note that in general, it is always better to decompose the covariates because the three different components may have different regression coefficients. However, in some cases due to small data or the nature of the variables, statistical significance for the additional covariates might not be attainable. That is, there is no statistically significant variation across the two higher levels and the added predictors might be too close to constant and may practically lose their usefulness. In principle, using the latent variable decomposition (1) is preferable to observed centering for covariates, i.e., converting the covariate to a dependent variable is preferable. Such decomposition prevents biases that may be introduced in the model due to not accounting for the measurement error in the mean/averages. However, observed centering might be preferable because no assumption on the distribution of the covariates is made. Latent centering, i.e., latent decomposition as in (1), assumes normally distributed components. If the distribution of the covariate deviates from a normal distribution substantially, observed centering decomposition of the covariate might be preferable.

The vectors $Y_{2,ij}$ and $Y_{3,j}$ can include not just the latent decomposition parts of the variables Y_{ijt} but can also include observed variables that are subject-specific or cluster-specific. Thus dependent variables observed at Level 2 and Level 3 can also be included in the model.

An additional extension of the above model is the possibility to include random slopes. All structural coefficients in equations (2) and (3) can be random effects, i.e., coefficients that vary over individuals and clusters. As

in the Mplus 3-level SEM framework, such coefficients are treated as latent variables on the higher levels and become a part of the vector $\eta_{2,ij}$ and $\eta_{3,j}$.

An essential part of the DSEM3 model is that all variable components on the within level are centered. Time-invariant, subject-specific or cluster-specific parts of the variables have been removed from the observed value when the variables or their lagged versions are used for the time-series modeling on the within level. If the variables are not centered, we can expect biased estimates, Hamaker and Grasman (2015). If the lagged variables are centered with the observed centering method, instead of the latent centering used in DSEM3, Nickell's bias occurs, see Nickell (1981) and Hamaker and Grasman (2015). Similarly, if a non-lagged variable is not centered, biases related to the big-fish-small-pond effect can occur, Marsh et al. (2014). Similarly, if a non-lagged variable is observed-centered, biases occur related to not accounting for the error in the average, Lüdtke et al. (2008). The concept of latent-centering is traditionally used for multilevel SEM, Muthén and Satorra (1995). More recently, it has been fully adopted in DSEM2, Asparouhov et al. (2018), as well as two-level SEM with random slopes Asparouhov and Muthén (2019).

The model description does not include a time-series model for the middle level. Such a time-series model can become of interest for those situations where the highest level cluster is the individual and the middle level clustering is a time-related variable such as bursts or day. A limited DSEM3 model estimation with time-series on both (lowest and middle) levels has been discussed in Asparouhov and Muthén (2024a). It should be noted that even when the middle level clustering is a time variable, before a time-series model can be established on both levels, a certain amount of data prerequisites will be needed: more than 5 observations at the lower level, more than 5 observations at the middle level, substantial percentage of variance at each of the two levels, i.e., greater than 2% of the total variance, autocorrelations that are statistically different from 0 and 1 on both levels. In addition, the data on the middle level must be in close proximity of time. Bursts of ILD that are several months apart are likely to have zero autocorrelation. Overall, there are 4 components competing to explain correlations among the observations: random effects $Y_{3,j}$, $Y_{2,ij}$, autocorrelation on the within level and autocorrelation on the middle level. Naturally the most difficult to establish will always be the auto-correlation on the middle level. Because of the high level of competition for correlation modeling among the four features, we expect that a very large amount of data (probably not what is currently

available in practical applications) would be required to establish the time series at the middle level. Currently, exploring such middle level autocorrelation can be done using the sample autocorrelation for the factor score estimates of $Y_{2,ij}$, or alternatively using a DSEM2 model for $Y_{2,ij}+Y_{3,j}$ based on their factor score estimates. Both of these ideas can be realized with a 2-step estimation where DSEM3 is the first step and DSEM2 is possibly the second step.

3 Model estimation

Here we do not provide a full description of the model estimation but rather limit the discussion to the strategy used for the estimation. It is largely based on the DSEM2 model estimation. The strategy boils down to using MCMC Bayesian estimation where the random effects at level 2 and level 3 are updated in separate Gibbs sampler steps. Conditional on the level 3 effects $Y_{3,j}$, the model becomes a DSEM2 model and thus $Y_{2,ij}$ can be updated as in DSEM2. Here the clustering of the observations is based on the middle level clustering variable. Similarly, conditional on the level 2 effects $Y_{2,ij}$ the model becomes a DSEM2 model and thus $Y_{3,j}$ can be updated as in DSEM2. Here the clustering of the observations is based on the highest level clustering variable.

The above strategy is simple but the simplicity might be costly. Updating parameters, random effects, missing data in many more separate blocks creates the possibility for non-convergence due to highly posteriorly correlated components being updated in different steps. This can cause slow convergence and non-convergence. The effects $Y_{3,j}$ and $Y_{2,ij}$ are likely to be highly posteriorly correlated. The fewer middle level units there are, the more posteriorly correlated the effects will be and if there is just one unit, that posterior correlation is 1. The exact same applies to random regression slopes present on both levels. Consider a simple example where the current estimate for these random effects are $Y_{3,j} + h$ and $Y_{2,ij} - h$, where $Y_{3,j}$ and $Y_{2,ij}$ are the true values and h is a large value compared to their posterior variance. We would expect that the posterior updating of the effect will eliminate h after several MCMC iterations. However, there is nothing in the data that helps in that process when the data components are updated separately. If the middle effect is conditioned upon and is set at $Y_{2,ij} - h$ then the level 3 effect must be near $Y_{3,j} + h$. The opposite updating also will not be able to remove

h. In fact the only evidence against h comes from the fact that the mean of $Y_{2,ij}$ is zero and the variance on both levels is to be as small as possible. Such evidence however is quite small compared to the weight of the within level model which may contain a large number of observations.

The question then arises as to whether there is a more efficient estimation algorithm that would update $Y_{3,j}$ and $Y_{2,ij}$ simultaneously. Clearly the answer is yes. As in DSEM2, the model can be reformulated to use the uncentered variables while the random effects are linearly transformed based on the autocorrelation. Thus the model becomes similar to a 3-level model where compact and simultaneous updating exists: $Y_{3,j}$ is updated first, then $[Y_{2,ij}|Y_{3,j}]$ is updated. A linear transformation of these random effects will produce the effects for DSEM3. Implementing such a more efficient algorithm is somewhat more intricate than what is currently implemented in Mplus. As we continue to gain experience with DSEM3 and the need arises for more complex models to be estimated, improved and more efficient algorithms are likely to be implemented in future versions of Mplus.

Bayesian estimation for latent centering of mediators with random slopes also provides an additional level of complexity. As described in Asparouhov and Muthén (2019), such modeling necessitates an additional split in the random effect updating: random intercepts and random slopes must be updated in 2 separate steps. Otherwise the posterior distribution is not explicit/normal. When this is added to the mix of DSEM3, we obtain 4 blocks of random effects that are updated separately. In these circumstances, the currently implemented estimation algorithm in Mplus faces more perilous waters to navigate and convergence issues must be strictly monitored. One exception/advancement is currently available in Mplus for non-DSEM3 3-level models with random slopes for latent centered mediators. For non-DSEM3 models the updating is based on the standard 3-level SEM updating and it uses just 2 blocks of random effects: random intercepts and random slopes, but across levels these are updated simultaneously.

4 Simulation studies

In the next few sections we present some key simple models that illustrate the basic concepts of DSEM3. The purpose of these simulations is to demonstrate that the estimation method works correctly, i.e., that the parameter estimates are unbiased and that the credibility intervals include the true pa-

rameter values approximately 95% of the time. For this illustration, we use the following models: a simple model with just one variable, a model with random-autoregressive coefficient, a regression model between two variables with random slope where both variables are autocorrelated, a vector autoregressive model with two variables, and a factor analysis model where the factor is autocorrelated.

4.1 Single variable model

Consider the single variable DSEM3 model

$$Y_{ijt} = Y_{1,ijt} + Y_{2,ij} + Y_{3,j}, \tag{8}$$

$$Y_{1,ijt} = \rho Y_{1,ijt-1} + \varepsilon_{ijt} \tag{9}$$

$$\varepsilon_{ijt} \sim N(0, \sigma_1), Y_{2,ij} \sim N(0, \sigma_2), Y_{3,j} \sim N(\nu, \sigma_3). \tag{10}$$

The model can also be presented as a hierarchical model as follows

$$\begin{aligned}
&\text{Level 1:} \quad Y_{ijt} = Y_{3,j} + Y_{2,ij} + \rho(Y_{ijt-1} - Y_{3,j} - Y_{2,ij}) + \varepsilon_{ijt} \\
&\text{Level 2:} \quad Y_{2,ij} = \varepsilon_{ij} \\
&\text{Level 3:} \quad Y_{3,j} = \nu + \varepsilon_j \\
&\varepsilon_{ijt} \sim N(0, \sigma_1), \varepsilon_{ij} \sim N(0, \sigma_2), \varepsilon_j \sim N(0, \sigma_3).
\end{aligned}$$

The third term in the level 1 equation reflects the centering of the predictor, that is, the lagged outcome variable. The level-3 mean of the predictor as well as the level-2 specific mean adjustment are subtracted to reduce the predictor to only the moment specific deviation that is relevant as a predictor for that moment. Note also that this is latent centering because we do not use the observed average values to center the predictor but we use the model variables which are latent variables representing the means.

The model has a total of 5 parameters: the three variances at each of the three levels, the intercept and the autocorrelation. Figure 1 shows the input file for a simple simulation study for this model. Here we use 50 level 3 units, each with 10 level 2 unit, and each of those contains 10 observations equally spaced. The results based on 100 replications are given in Figure 2 and show that the estimation performs well.

There are 3 different models that such data could be analyzed with in the absence of the DSEM3 estimation. The first is just a regular 3-level model

that ignores the autocorrelation parameter. The model can be estimated with the maximum-likelihood estimator or with the Bayesian estimator and in most situations the results are nearly identical. Here we use the Bayesian estimator to avoid any discrepancy due to the type of estimator. The second method is a DSEM2-L2 model where the third level of clustering is ignored. Such a method would be applicable when the level 2 cluster variable is individual and the level 3 cluster variable is a nesting structure where individuals belong to, such as classrooms. The third method is DSEM2-L3 where the middle level clustering is ignored and the data within each level 3 unit is treated as an equally spaced set of sequential observations, a total of 100 in each level 3 unit. This method would be applicable for situations where the middle level clustering is a time variable.

The results for all these models are given in Table 1 and remarkably all the results are as expected. The 3-level model shows bias in σ_2. Not accounting for the autocorrelation will inflate the middle level intercept which can compensate somewhat for the autocorrelation. Underestimation is visible also in the standard error of σ_1 which is caused by ignoring the non-independence on the within level. For the DSEM2-L2 model, σ_2 estimate matches $\sigma_2+\sigma_3$ value. The mean and variance standard errors are underestimated due to ignoring the higher level clustering and the non-independence of the level 2 units. For the DSEM2-L3 model, the autocorrelation is inflated as an attempt to explain the missing level 2 effect. The σ_1 parameter is also inflated because not all the level 2 effect caused correlation can be fitted via autocorrelation. All other parameters remain intact. For example, in the 3-level model, σ_3 did not show any bias. That is, not accounting for the autocorrelation appears to strictly shift the 2-level variance only, but not the third level variance. In DSEM-L2, not accounting for the third level clustering did not affect the autocorrelation. In DSEM2-L3, ignoring the middle level clustering resulted in overestimation in the within level model but the third level variance is intact. Understanding the dynamics between these four models can be quite useful in practice. Estimating DSEM3 should be accompanied by these more restricted model estimations.

Note here that the results in Table 1 also point to a small bias in σ_3 for all models, although coverage is not affected. Note, however, that this is a well understood issue with the Bayesian estimator. We will discuss further various effects on the sample size at each level further below. For now we provide the following explanation. The quality of the estimates of the level 3 parameters are driven entirely by the number of level 3 units. The higher

the number of level 3 units is, the closer we are to the asymptotic behavior of consistency for likelihood based estimation. When the number of level 3 units is smaller, some finite sample size bias is expected. In addition, introducing a proper weakly informative prior for that parameter will likely influence the estimates and can be adjusted to reduce the bias. In contrast, the level 2 units and level 1 units are usually sufficiently large so that the asymptotic behavior is in effect. For our example, the number of level 2 units is 500 and the number of level 1 units is 5000 and therefore we can expect that the estimates will be consistent and that weakly informative priors will not have any effect on the estimates.

For the univariate model given above the variance of Y_{ijt} is decomposed 4-ways:

$$Var(Y_{ijt}) = \sigma_1 + \sigma_1 \frac{\rho^2}{1-\rho^2} + \sigma_2 + \sigma_3. \tag{11}$$

The first term is the unexplained variance, the second term is the variance explained by the autocorrelation, the third and fourth terms are the effects of the second and third level clustering. We can think of this also in terms of intra-class correlation, i.e., in terms of the proportion each of the modeling components represents $ICC_1 = \sigma_1/Var(Y_{ijt})$, $ICC_{AR} = \sigma_1\rho^2/(Var(Y_{ijt})(1-\rho^2))$, $ICC_2 = \sigma_2/Var(Y_{ijt})$, and $ICC_3 = \sigma_3/Var(Y_{ijt})$. In our simulated example the 4-way variance decomposition amounts to this: $ICC_1 = 53\%$, $ICC_{AR} = 5\%$, $ICC_2 = 16\%$, $ICC_3 = 26\%$. This variance decomposition has two implications. First, it allows us to focus on and understand where the variance for a variable is coming from. Second, for a variable to be truly a DSEM3 variable, all 4 components in this decomposition should be non-zero and also be practically and meaningfully non-zero. We argue here that components that are less than 1 or 2 percents of the total variance are probably not of practical importance in most situations. Components that account for less than 1 or 2 percentage points from the total variance can be removed and a simpler model such as: 3-level, DSEM2-L2 or DSEM2-L3 should be considered as the base to build upon.

Figure 1: Three-level DSEM single variable simulation study

```
montecarlo:
                names are y;
                nobservations = 5000;
                nreps = 100;
                csizes = 50[10(10)];
                ncsize = 1[1];
                lagged=y(1);

ANALYSIS:  type = threelevel; estimator=bayes;
           process=2;

model population:

                %within%
                y*1;
                y on y&1*0.3;

                %between LEVEL2%
                y*.3;

                %between LEVEL3%
                y*.5;
                [y*1];

model:
                %within%
                y*1;
                y on y&1*0.3;

                %between LEVEL2%
                y*.3;

                %between LEVEL3%
                y*.5;
                [y*1];
```

Figure 2: Three-level DSEM single variable simulation study results

MODEL RESULTS

	Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Within Level							
Y ON							
Y&1	0.300	0.3074	0.0175	0.0179	0.0004	0.940	1.000
Residual Variances							
Y	1.000	1.0095	0.0215	0.0221	0.0005	0.960	1.000
Between LEVEL2 Level							
Variances							
Y	0.300	0.3041	0.0375	0.0353	0.0014	0.930	1.000
Between LEVEL3 Level							
Means							
Y	1.000	1.0030	0.1021	0.1094	0.0103	0.960	1.000
Variances							
Y	0.500	0.5357	0.1168	0.1300	0.0148	0.950	1.000

Table 1: Single variable DSEM3 model: estimate(coverage)

Parameter	True Value	DSEM3	3-level	DSEM2-L2	DSEM2-L3
σ_1	1	1.01(.96)	1.01(.84)	1.01(.93)	1.16(.00)
σ_2	0.3	.30(.93)	0.40(.20)	.80(.82)	-
σ_3	0.5	.54(.95)	0.56(.97)	-	.56(.94)
ρ	0.3	.31(.94)	-	.31(.98)	.40(.00)
ν	1	1.00(.96)	1.01(.96)	1.00(.60)	1.01(.98)

4.2 Single variable model with random autocorrelation

Here we consider the single variable DSEM3 model with random autocorrelation, i.e., with cluster specific autocorrelation

$$Y_{ijt} = Y_{1,ijt} + Y_{2,ij} + Y_{3,j}, \tag{12}$$

$$Y_{1,ijt} = \rho_{ij} Y_{1,ijt-1} + \varepsilon_{ijt} \tag{13}$$

$$\rho_{ij} = \rho_{2,ij} + \rho_{3,j} \tag{14}$$

$$\varepsilon_{ijt} \sim N(0, \sigma_1), Y_{2,ij} \sim N(0, \sigma_2), Y_{3,j} \sim N(\nu, \sigma_3). \tag{15}$$

$$\rho_{2,ij} \sim N(0, v_2), \rho_{3,j} \sim N(\rho, v_3) \tag{16}$$

$$Cov(\rho_{2,ij}, Y_{2,ij}) = c_2, Cov(\rho_{3,j}, Y_{3,j}) = c_3. \tag{17}$$

The model can also be presented as a hierarchical model as follows

$$\begin{aligned}
&\text{Level 1:} && Y_{ijt} = Y_{3,j} + Y_{2,ij} + \rho_{ij}(Y_{ijt-1} - Y_{3,j} - Y_{2,ij}) + \varepsilon_{ijt} \\
&\text{Level 2:} && Y_{2,ij} = \varepsilon_{ij} \\
& && \rho_{ij} = \rho_{3,j} + \delta_{ij} \\
&\text{Level 3:} && Y_{3,j} = \nu + \varepsilon_j \\
& && \rho_{3,j} = \rho + \delta_j
\end{aligned}$$

$$\varepsilon_{ijt} \sim N(0, \sigma_1), \varepsilon_{ij} \sim N(0, \sigma_2), \varepsilon_j \sim N(0, \sigma_3), \delta_{ij} \sim N(0, v_2), \delta_j \sim N(0, v_3)$$

$$Cov(\varepsilon_{ij}, \delta_{ij}) = c_2, Cov(\varepsilon_j, \delta_j) = c_3.$$

In this model the autocorrelation varies not just over level 3 units but also within level 3 units and is level 2 cluster specific. Figure 3 shows the Mplus

input for conducting a simulation study for the above model and Figure 4 shows the results of the simulation. The results indicate that the estimation performs well.

In this simulation study we have also introduced missing data. We doubled the size of the level 2 units to 20 here but each observation is missing with probability of 50%, so on average there are 10 observations within each level 2 cluster. This type of missing data is similar to the missing data created by the TINTERVAL command used when the observations are irregularly spaced. The TINTERVAL command in DSEM3 works the same way as for DSEM2. When observations are spaced further apart, missing data rows are inserted so that the actual data that is analyzed more fully represents the varying distances between the observations.

The next two sections describe the most basic bivariate models.

Figure 3: Three-level DSEM single variable with random autocorrelation simulation study

```
montecarlo:
                names are y;
                nobservations = 10000;
                nreps = 100;
                csizes = 50[10(20)];
                ncsize = 1[1];
                lagged=y(1);
                missing=y;

model missing: [y*0];


ANALYSIS:  type = threelevel random; estimator=bayes;
           process=2;

model population:

                %within%
                y*1;
                s | y on y&1;

                %between LEVEL2%
                y*.3; s*0.01;
                y with s*0.02;

                %between LEVEL3%
                y*.5; s*0.01;
                [y*1]; [s*0.3];
                y with s*0.02;

model:
                %within%
                y*1;
                s | y on y&1;

                %between LEVEL2%
                y*.3; s*0.01;
                y with s*0.02;

                %between LEVEL3%
                y*.5; s*0.01;
                [y*1]; [s*0.3];
                y with s*0.02;
```

Figure 4: Three-level DSEM single variable with random autocorrelation simulation study results

MODEL RESULTS

		ESTIMATES		S. E.	M. S. E.	95%	% Sig
	Population	Average	Std. Dev.	Average		Cover	Coeff
Within Level							
Residual Variances							
Y	1.000	1.0018	0.0217	0.0219	0.0005	0.940	1.000
Between LEVEL2 Level							
Y WITH							
S	0.020	0.0170	0.0113	0.0119	0.0001	0.970	0.310
Variances							
Y	0.300	0.3067	0.0360	0.0332	0.0013	0.900	1.000
S	0.010	0.0125	0.0047	0.0051	0.0000	0.880	1.000
Between LEVEL3 Level							
Y WITH							
S	0.020	0.0206	0.0196	0.0224	0.0004	0.950	0.170
Means							
Y	1.000	1.0082	0.0945	0.1120	0.0089	0.970	1.000
S	0.300	0.3004	0.0272	0.0265	0.0007	0.920	1.000
Variances							
Y	0.500	0.5622	0.1210	0.1384	0.0184	0.940	1.000
S	0.010	0.0132	0.0062	0.0067	0.0000	0.940	1.000

4.3 The DSEM3 regression model

This model features three-level modeling, latent centering for the endogenous and the exogenous variables, autocorrelation on the within level for both the endogenous and the exogenous variables, as well as a random regression slope that varies over both level 2 and level 3 clusters. The model can be described more formally as follows

$$Y_{ijt} = Y_{1,ijt} + Y_{2,ij} + Y_{3,j}$$

$$X_{ijt} = X_{1,ijt} + X_{2,ij} + X_{3,j}$$

$$Y_{1,ijt} = \rho_1 Y_{1,ijt-1} + s_{ij} X_{1,ijt} + \varepsilon_{ijt}$$

$$X_{1,ijt} = \rho_2 X_{1,ijt-1} + \xi_{ijt}$$

$$s_{ij} = s_{2,ij} + s_{3,j}$$

All the variables in the above equations are assumed normally distributed. The three variables on level 2: $Y_{2,ij}$, $X_{2,ij}$, and $s_{2,ij}$ can form any kind of structural model at level 2, including regressions and correlation models. The three variables on level 3: $Y_{3,j}$, $X_{3,j}$, and $s_{3,j}$ can form any kind of structural model at level 3. Figure 5 shows an Mplus simulation study example of the above model and Figure 6 shows the results. We see here that the bias is minimal and the coverage is near the nominal levels for all parameters. Slightly larger bias is observed for the parameter estimates at level 3, however, this bias is the results of the small number of level 3 units (in this simulation it is 50) and the bias is expected to gradually disappear as the number of units increases and the asymptotic guarantees can kick in. Alternatively, using weakly informative proper priors for the parameters at level 3 can be used to reduce the bias when the number of level 3 units is small. Mplus uses by default the improper prior of a constant over the entire range for each parameter.

Figure 5: Three-level DSEM regression

```
montecarlo:
                names are y x;
                nobservations = 10000;
                nreps = 100;
                CSIZES = 50[10(20)];
                ncsize = 1[1];
                lagged=y(1) x(1);

ANALYSIS:  TYPE = threelevel random;
process=2; estimator=bayes;

model population:
                %within%
                y*1.2 x*0.8;
                s | y on x;
                y on y&1*0.3;
                x on x&1*0.5;

                %between LEVEL2%
                y*1.5 x*0.5 s*0.02;
                y with s*0.1;
                y on x*0.4;

                %between LEVEL3%
                y*1.3; s*0.2; x*0.5;
                [y*2.1 s*0.3 x*1];
                y with s*0.2;
                y on x*-0.5;

model:
                %within%
                y*1.2 x*0.8;
                s | y on x;
                y on y&1*0.3;
                x on x&1*0.5;

                %between LEVEL2%
                y*1.5 x*0.5 s*0.02;
                y with s*0.1;
                y on x*0.4;

                %between LEVEL3%
                y*1.3; s*0.2; x*0.5;
                [y*2.1 s*0.3 x*1];
                y with s*0.2;
                y on x*-0.5;
```

Figure 6: Three-level DSEM regression results

```
MODEL RESULTS

                    ESTIMATES             S. E.     M. S. E.  95%  % Sig
               Population   Average   Std. Dev.   Average             Cover Coeff

Within Level

 Y          ON
  Y&1              0.300     0.3010     0.0094     0.0095     0.0001 0.960 1.000

 X          ON
  X&1              0.500     0.5074     0.0101     0.0105     0.0002 0.900 1.000

 Residual Variances
  Y                1.200     1.2056     0.0187     0.0179     0.0004 0.900 1.000
  X                0.800     0.8039     0.0102     0.0119     0.0001 0.950 1.000

Between LEVEL2 Level

 Y          ON
  X                0.400     0.3989     0.0914     0.0963     0.0083 0.930 1.000

 Y        WITH
  S                0.100     0.0978     0.0178     0.0187     0.0003 0.930 1.000

 Variances
  X                0.500     0.5057     0.0444     0.0450     0.0020 0.910 1.000
  S                0.020     0.0204     0.0050     0.0047     0.0000 0.910 1.000

 Residual Variances
  Y                1.500     1.5479     0.1231     0.1185     0.0173 0.960 1.000

Between LEVEL3 Level

 Y          ON
  X               -0.500    -0.5827     0.2770     0.2620     0.0828 0.900 0.580

 Y        WITH
  S                0.200     0.2402     0.0998     0.1124     0.0115 0.940 0.790

 Means
  X                1.000     1.0014     0.1249     0.1114     0.0154 0.910 1.000
  S                0.300     0.3019     0.0700     0.0708     0.0049 0.950 0.980

 Intercepts
  Y                2.100     2.2040     0.3383     0.3228     0.1241 0.900 1.000

 Variances
  X                0.500     0.5400     0.1115     0.1357     0.0139 0.960 1.000
  S                0.200     0.2351     0.0406     0.0557     0.0029 0.950 1.000
```

4.4 The DSEM3 vector autoregressive (VAR) model

The next basic bivariate model is the VAR model where the autocorrelation is bivariate and the variables from the current period are predicted by both variables from the prior period. The model is described by the following equations

$$Y_{ijt} = Y_{1,ijt} + Y_{2,ij} + Y_{3,j}$$
$$Z_{ijt} = Z_{1,ijt} + Z_{2,ij} + Z_{3,j}$$
$$Y_{1,ijt} = r_1 Y_{1,ij,t-1} + r_2 Z_{1,ij,t-1} + \varepsilon_{1ijt}$$
$$Z_{1,ijt} = r_3 Y_{1,ij,t-1} + r_4 Z_{1,ij,t-1} + \varepsilon_{2ijt}.$$

The pairs of variables ε_{1ijt} and ε_{2ijt}; $Y_{2,ij}$ and $Z_{2,ij}$; $Y_{3,j}$ and $Z_{3,j}$; are normally distributed correlated variables and form the models on each of the three levels. Figure 7 shows an example of an Mplus simulation study for the above model and Figure 8 shows the results. The bias is minimal and the coverage is near the nominal levels for all parameters. Figure 8 also shows the quality of the estimation of the random effects/intercepts $Y_{2,ij}, Y_{3,j}, Z_{2,ij}, Z_{3,j}$. Currently this output shows the correlation between the random effect estimate for each cluster and their true value as a function of the cluster. These correlations are averaged across the 100 replications in the simulation study. The MSE is also included. In general, here we want to see high values for the correlation between the estimates and their true values. However, there is no specific guidance for how high the correlations should be as these will typically depend on the model, sample size and parameter values. The correlations can mostly be used to evaluate how easy it is to estimate a particular random effect as compared to another. From the results of Figure 8 we conclude that the level 3 random intercepts are much easier to estimate than the level 2 random intercepts. This is as expected because the level 3 effects are measured by 200 values (number of observations in level 3 units) while the level 2 effects are measured by 20 values only. In this particular simulation study each replication takes 0.3 seconds.

Figure 7: Three-level DSEM VAR

```
montecarlo:
                names are y z;
                nobservations = 10000;
                nreps = 100;
                csizes = 50[10(20)];
                ncsize = 1[1];
                lagged=y-z(1);

analysis:       type= threelevel;
                estimator=bayes; proc=2;

model population:

                %within%
                y-z*1.2;
                y on y&1*0.4 z&1*0.1;
                z on y&1*0.2 z&1*0.3;
                y with z*0.2;

                %between LEVEL2%
                y-z*.4; y with z*0.2;

                %between LEVEL3%
                y-z*.5; y with z*0.3;
                [y*2 z*1];

model:
                %within%
                y-z*1.2;
                y on y&1*0.4 z&1*0.1;
                z on y&1*0.2 z&1*0.3;
                y with z*0.2;

                %between LEVEL2%
                y-z*.4; y with z*0.2;

                %between LEVEL3%
                y-z*.5; y with z*0.3;
                [y*2 z*1];
```

Figure 8: Three-level DSEM VAR results

MODEL RESULTS

		ESTIMATES		S. E.	M. S. E.	95%	% Sig
	Population	Average	Std. Dev.	Average		Cover	Coeff
Within Level							
Y ON							
Y&1	0.400	0.4042	0.0104	0.0108	0.0001	0.960	1.000
Z&1	0.100	0.1034	0.0099	0.0105	0.0001	0.960	1.000
Z ON							
Y&1	0.200	0.2037	0.0110	0.0103	0.0001	0.900	1.000
Z&1	0.300	0.3044	0.0102	0.0108	0.0001	0.950	1.000
Y WITH							
Z	0.200	0.2031	0.0134	0.0128	0.0002	0.960	1.000
Residual Variances							
Y	1.200	1.2048	0.0203	0.0179	0.0004	0.910	1.000
Z	1.200	1.2045	0.0181	0.0177	0.0003	0.930	1.000
Between LEVEL2 Level							
Y WITH							
Z	0.200	0.2065	0.0280	0.0318	0.0008	0.990	1.000
Variances							
Y	0.400	0.4074	0.0360	0.0413	0.0013	0.980	1.000
Z	0.400	0.4103	0.0394	0.0391	0.0016	0.970	1.000
Between LEVEL3 Level							
Y WITH							
Z	0.300	0.3335	0.1078	0.1139	0.0126	0.940	0.980
Means							
Y	2.000	1.9976	0.1078	0.1131	0.0115	0.960	1.000
Z	1.000	0.9935	0.1136	0.1107	0.0128	0.940	1.000
Variances							
Y	0.500	0.5681	0.1178	0.1420	0.0184	0.950	1.000
Z	0.500	0.5552	0.1313	0.1383	0.0201	0.920	1.000

CORRELATIONS AND MEAN SQUARE ERROR OF THE TRUE FACTOR VALUES AND THE FACTOR SCORES

	CORRELATIONS		MEAN SQUARE ERROR	
	Average	Std. Dev.	Average	Std. Dev.
B2_Y	0.788	0.016	0.390	0.013
B2_Z	0.810	0.016	0.371	0.014
B3_Y	0.945	0.016	0.230	0.024
B3_Z	0.946	0.015	0.225	0.023

4.5 The DSEM3 factor analysis model

Suppose that a latent factor is measured by p indicators across multiple time points t for multiple individuals i in multiple clusters j. The DSEM3 factor analysis model is given by the following equations

$$Y_{pijt} = Y_{1,pijt} + Y_{2,pij} + Y_{3,pj}$$

$$Y_{1,pijt} = \lambda_{1p}\eta_{ijt} + \varepsilon_{1,pijt}$$

$$\eta_{ijt} = r\eta_{ij,t-1} + \xi_{ijt}$$

$$Y_{2,pij} = \lambda_{2p}\eta_{ij} + \varepsilon_{2,pij}$$

$$Y_{3,pj} = \nu_p + \lambda_{3p}\eta_j + \varepsilon_{3,pj}.$$

Here η_{ijt}, η_{ij} and η_j are the factors on the three different levels. The loadings are free and the factor variance/residual variance is fixed to 1 for identification purposes. Without the autoregression for the level 1 factor, the model is simply a 3-level factor analysis. Without the third level model, the model would be a DSEM2 factor model. Note that given the 4 indices above, the model is essentially a 4 level model, where the new 4-th level is essentially the multivariate index p. If needed, the loadings can be held equal across the levels and the factor variance can be estimated at level-2 and level-3. This will yield a latent-centering decomposition of the factor variable which mimics the observed variables decomposition by levels.

Figure 9 shows an Mplus simulation study example of the above model and Figure 10 shows the results for a selection of the parameters. The bias is minimal and the coverage is near the nominal levels for all parameters.

Figure 9: Three-level DSEM factor analysis

```
montecarlo:
                names are y1-y5;
                nobservations = 20000;
                nreps = 100;
                csizes = 50[20(20)];
                ncsize = 1[1];

analysis: type = threelevel;
                  estimator=bayes; proc=2;

model population:

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

        %between LEVEL2%
        f2 by y1-y5*0.4;
        f2@1 y1-y5*0.2;

        %between LEVEL3%
        f3 by y1-y5*0.6;
        f3@1 y1-y5*0.2;

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

        %between LEVEL2%
        f2 by y1-y5*0.4;
        f2@1 y1-y5*0.2;

        %between LEVEL3%
        f3 by y1-y5*0.6;
        f3@1 y1-y5*0.2;
```

Figure 10: Three-level DSEM factor analysis

	Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Within Level							
F BY							
Y1	1.000	0.9999	0.0098	0.0094	0.0001	0.920	1.000
Y5	1.000	0.9999	0.0103	0.0093	0.0001	0.910	1.000
F ON							
F&1	0.400	0.4002	0.0085	0.0087	0.0001	0.940	1.000
Residual Variances							
Y1	1.000	1.0026	0.0120	0.0128	0.0001	0.980	1.000
Y5	1.000	1.0010	0.0124	0.0128	0.0002	0.970	1.000
F	1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
Between LEVEL2 Level							
F2 BY							
Y1	0.400	0.3994	0.0348	0.0296	0.0012	0.900	1.000
Y5	0.400	0.4018	0.0289	0.0292	0.0008	0.970	1.000
Variances							
F2	1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
Residual Variances							
Y1	0.200	0.1998	0.0168	0.0145	0.0003	0.910	1.000
Y5	0.200	0.1981	0.0141	0.0145	0.0002	0.950	1.000
Between LEVEL3 Level							
F3 BY							
Y1	0.600	0.6376	0.1042	0.1088	0.0122	0.920	1.000
Y5	0.600	0.6378	0.1120	0.1097	0.0138	0.890	1.000
Intercepts							
Y1	0.000	-0.0142	0.1011	0.1132	0.0103	0.970	0.030
Y5	0.000	-0.0144	0.1095	0.1132	0.0121	0.960	0.040
Variances							
F3	1.000	1.0000	0.0000	0.0000	0.0000	1.000	0.000
Residual Variances							
Y1	0.200	0.2121	0.0563	0.0637	0.0033	0.950	1.000
Y5	0.200	0.2112	0.0539	0.0629	0.0030	0.990	1.000

5 Data requirements

A three level DSEM model would ideally be applied to the analysis of data that is reasonable in size at each level, meaning sufficiently large. That may be a problem for many DSEM3 applications as often data is collected without a specific model in mind. Human data is also sometimes too expensive to collect in unlimited quantities. Thus, the DSEM3 methodology must in principle be evaluated for situations with insufficient data. In this section we will perform some simulation studies to inform on some of the issues that arise. However, we can not provide a simple answer of this sort: a minimum of n_1 units are needed on level 1, a minimum of n_2 units are needed on level 2, and a minimum of n_3 units are needed on level 3. Such an approach would be naive. The data that is needed depends on the model. The more complex the model is, the more data is needed. Random/subject specific autocorrelation would naturally need longer time series than fixed/subject invariant autocorrelation. Furthermore, the amount of data that is needed depends on the parameter values. It is much easier to make a distinction between a random intercept and autocorrelation if the autocorrelation is not very high.

The simulation studies that we consider here are based on balanced designs where each level-3 unit has the same number of level-2 units which are also of the same size. In practice such assumptions are rarely met and we use the average size of level-2 units and the average number of level-2 units within each level-3 unit as measures of data size. We assume that the performance of the methodology for the unbalanced design would be similar to the case of balanced design with the same average values. This assumption, however, applies mostly to simpler models. More complex models such as those that have many cluster specific parameters may have substantially worse performance with unbalanced designs. Consider for example the situation where the data contains many level-2 units of size 1. If the three levels represent within-day, between-day, and between-person variation, units of size 1 would mean having many cases with only 1 observation per day. Suppose also that the model contains more than one level-2 specific random effect. Those effects are unidentified by the data they are intended to model in the clusters of size 1 and can only be estimated from the general information from all other clusters. If the number of level-2 units of size 1 is substantial, that information would also appear practically unidentified as it contains many unidentified effects. This would result in poor performance, slow convergence

or non-convergence. If the design is balanced or if the model is simpler such issues would not arise. We conclude that in practical applications it is important to consider not just the average cluster sizes but also the distribution of the cluster sizes. In particular, if there are many very small level-2 units, random slopes and random auto-regressive coefficients should not be level-2 specific but only level-3 specific. Preference should be given to models where only the random intercepts are level-2 specific. Mplus currently does not provide a detailed level-3 and level-2 joint cluster size distribution. To evaluate the cluster size distribution it would be necessary to estimate a two-level model where the clustering variable is the level-2 clustering variable. Mplus will produce the distribution of the level-2 cluster sizes. The number of level-2 units in each level-3 unit should also be considered. If many level-3 units contain just one level-2 unit similar non-identification may occur.

In multilevel models, 3-level SEM as well as DSEM2, a random effect parameter is justifiable only when it can replace at least 5 or 10 fixed parameters. For example, if on level 3 we have just two units, it is more reasonable to use a 2-group 2-level analysis, i.e., replace the random intercept with two fixed intercept parameters: one for each group. Note that 2-group 2-level model still has many more parameters. A simplified version of using the grouping variable as a covariate, essentially results in a 2-group analysis where only the intercept parameters are group specific. Thus with a small number of units at the highest level, three level models are not recommended and instead two-level models should be used. To some extent this applies also to the situation with very few level 2 units. There are two or three alternative modeling approaches to consider. One approach is to use a two-level model where the clustering is based only on the level 2 clustering variable (which essentially combines the two clustering variables). A second approach which is feasible in situations where the second level clustering is based on identical information across the three levels: for example (men in level 3 units v.s. women in level 3 units or first and second bursts of ILD). In this case again a covariate can be used instead of a clustering variable. Alternatively a multivariate approach can be used where the data is set in a wide format with each of the level 2 variables represented as parallel processes, see Hamaker et al. (2023). Finally, if the number of level 1 units is small, a three level model can be reformulated as a two-level wide model. It is important to understand that a time series model with very few observations in the sequence is generally difficult since the autocorrelation would be difficult to separate from the correlation implied by the random intercept.

Table 2: Small number of units: absolute bias(coverage)

N1/N2/N3	10/10/50	5/10/50	10/5/50	10/10/5
σ_1	.01(.96)	.05(.70)	.01(.94)	.03(.92)
σ_2	.00(.93)	.06(.84)	.01(.94)	.03(.91)
σ_3	.04(.95)	.03(.95)	.00(.97)	1.33(.87)
ρ	.01(.94)	.06(.66)	.00(.99)	.01(.94)
ν	.00(.96)	.01(.90)	.00(.96)	.02(1.00)

5.1 Simulation studies with small number of units at each level

In this section we explore the effect of having a small number of units at each of the three levels. We will utilize as a starting point the simulation study described in Section 4.1. For simplicity, in our simulations, all units on the third level have the same number of level 2 units. We denote that number by N_2. Also, each level two unit has the same number of observations. We denote that by N_1. We also denote the number of level 3 units by N_3. Thus the total sample size is $N = N_1N_2N_3$. We compare the setup in Section 4.1 to the setup where each of these 3 numbers is lowered to 5. The results are presented in Table 2. When N_3 is lowered to 5 a large bias is seen for σ_3, but this is expected. When N_2 is lowered to 5, the estimates remain good. When N_1 is lowered to 5, small bias can be seen for σ_1, σ_2 and ρ. As previously discussed, when N_1 gets smaller an identification issue arises between σ_2 and ρ. Even though the bias is small, coverage for these parameters drops. Overall the performance in all 3 situations appears acceptable but not perfect.

Next we repeat the above stimulation with the autocorrelation parameter set to 0.8. The results are given in Table 3. In the 10/10/50 case, the estimates mostly remained good. A small bias in the autocorrelation appears to lead to lower coverage. In the other three cases the results are now worse than with autocorrelation of 0.3. Some non-convergence is also recorded. The σ_2 parameter estimates in particular appear to be poor and much worse now. One explanation is that when the autocorrelation is 0.8 and N_1 is 5 or 10, the implied autocorrelation between the most distant observations in the time series is not zero but $0.33 = .8^5$ or $0.11 = .8^{10}$. This means that the autocorrelation is non-zero for any two variables in the time series and such

Table 3: Small number of units and high autocorrelation: absolute bias(coverage)

N1/N2/N3	10/10/50	5/10/50	10/5/50	10/10/5
Convergence	100%	70%	100%	70%
σ_1	.01(.96)	.01(.93)	.01(.94)	.02(.93)
σ_2	.02(.90)	.47(.61)	.16(.91)	.49(.90)
σ_3	.02(.96)	.01(.98)	.02(.97)	2.08(.89)
ρ	.02(.62)	.03(.54)	.02(.82)	.00(.97)
ν	.01(.90)	.00(.95)	.00(.96)	.02(1.00)

correlation can be absorbed by the random intercept thereby compromising the estimation of σ_2. It is clear that the combination of a small number of units on any level and high autocorrelation leads to more challenging estimation.

5.2 Latent centering v.s. Observed centering

Lüdtke et al. (2008) shows that latent centering has an advantage over observed centering as it accounts for the error in the centering. This applies to two-level models as well as three level models with and without random slopes. Often however the differences are small and observed centering can be used as a simpler alternative model and as a replacement of latent centering model for situations where latent centering leads to convergence problems.

In this section we conduct a simulation study to compare the performance of latent and observed centering. Here we do not use a dynamic model but a simple 3-level model. Figure 11 shows the input file for generating the data and analyzing it with latent centering. Figure 12 shows the input file for analyzing the same data using observed centering. We conduct the simulation study with a smaller and a bigger sample size. For the smaller sample size case, we use $N_1 = 10, N_2 = 10, N_3 = 20$. For the bigger sample size situation, we use $N_1 = 20, N_2 = 20, N_3 = 50$.

Figures 13 and 14 show a comparison of results for the key regression parameters for the smaller and bigger samples. For the smaller sample size, we see that the advantages of latent centering are somewhat smaller. Depending on the particular parameter one is interested in and the choice of criterion: bias, coverage, or MSE, either one of the two estimation possibilities could be perceived as preferable. However, for the larger sample size, the advantages of the latent centering method are rather clear. The observed centering method yields biased parameter estimates for one of the parameters which also results in poor coverage and much larger MSE.

General guidance for the latent v.s. observed centering comparison is somewhat difficult to formulate as we have several levels of variability: the sample size on each of the three levels N_1, N_2 and N_3 as well as the model complexity. Nevertheless, general theory as well as the precise observed centering bias for two-level models estimated in Asparouhov and Muthén (2006), can be used to illuminate at least some scenarios. First, as N_3 increases, the latent centering is guaranteed to yield unbiased estimates, correct coverage and smaller MSE than observed centering. This is because larger N_3 is the only condition needed to claim the ML asymptotic properties of the latent centering method. The smaller the values of N_1 and N_2, the larger the observed centering bias will be as the sample averages will have larger errors. Larger values of N_3 will not improve the observed centering bias but larger values for N_1 and N_2 will. The observed centering bias depends on the model

parameter values as well as N_1 and N_2. The larger the difference between the regression coefficients on the three levels, the larger the observed centering bias will be. If N_1, N_2 and N_3 are all smaller, the observed centering may have smaller MSE, even if the parameters are biased. With smaller samples, i.e. small N_1, N_2 and N_3, the increased variability of the latent centering estimates, associated with the fact that the method is less parsimonious in terms of estimating more random effects, may overpower the observed centering bias and result in larger MSE. Furthermore, in the presence of missing data, observed centering becomes even less reliable because the averages may have an additional bias if the missing data is not missing completely at random (MCAR). Latent centering will not be impacted by missing data and is guaranteed to work well even when the data is missing at random (MAR), i.e., where missing values in one variable are associated with the values of another variable. Overall, the latent centering method is generally more reliable and should be the first choice estimator. Observed centering could be considered either as an alternative solution to latent centering converges problems if such arise, or as a more accurate solution in the small sample situation. Simulation studies might be necessary to justify observed centering use.

The above comparison between observed and latent centering applies equally well to DSEM3 models. However, as described in Section 3, in the DSEM3 framework, the latent centering with random slopes is based on a less efficient algorithm that is more likely to result in non-convergence. Therefore, observed centering might be needed more often for DSEM3 models. In fact, if observed centering is easily accessible, in the DSEM3 framework both models should be estimated and compared.

We conclude this section with one final clarification. The comparison described above is specific to the case when a predictor is multiplied by a random slope that varies over level 2 and/or level 3 units and the predictor is a variable decomposed on all three levels. It does not apply to the cases when the predictor is a lagged variable, when the slope is not random or when the predictor can not be decomposed (for example time). In these situations observed v.s. latent centering can still be formulated, however, the estimation methodology is different and only latent centering is used. If the slope is not random, then the latent centering estimation algorithm is quite efficient because all random effects (only random intercepts) are updated simultaneously. The lagged variables latent centering is also somewhat different as the same centering variable is applied to the lagged and the current variables.

This allows the model to be reformulated so that the centering is just one scaled variable. If the covariate is not decomposed on all three levels, then no centering is done at all.

Figure 11: Latent centering simulation study

```
montecarlo:
              names are y x;
              nobservations = 2000;
              nreps = 100;
              csizes = 20[10(10)];
              ncsize = 1[1];
              save=a*.dat;
              repsave=all;

analysis:  type = threelevel random;
process=2; estimator=bayes;

model population:
              %within%
              y*1.2 x*0.8;
              s | y on x;

              %between LEVEL2%
              y*0.3 s*0.2 x*0.5;
              y with s*0.1;
              y on x*0.4;
              s on x*0.1;

              %between LEVEL3%
              y*0.4; s*0.2; x*0.5;
              [y*2.1 s*1 x*1];
              y with s*0.2;
              y on x*-0.5;
              s on x*0.2;

model:
              %within%
              y*1.2 x*0.8;
              s | y on x;

              %between LEVEL2%
              y*0.3 s*0.2 x*0.5;
              y with s*0.1;
              y on x*0.4;
              s on x*0.1;

              %between LEVEL3%
              y*0.4; s*0.2; x*0.5;
              [y*2.1 s*1 x*1];
              y with s*0.2;
              y on x*-0.5;
              s on x*0.2;
```

Figure 12: Observed centering simulation study

```
variable:
        names are y x c2 c3;
        cluster=c3 c2;
        usevar = y x x2 x3;
        within=x;
        between=(c2) x2;
        between=(c3) x3;

        define:
        x3=cluster_mean(x c3);
        x2=cluster_mean(x c2);
        x2=x2-x3;
        x=x-x2-x3;

data: file=alist.dat; type=montecarlo;


analysis:  type = threelevel random;
process=2; estimator=bayes;

model:
                %within%
                y*1.2 x*0.8;
                s | y on x;

                %between c2%
                y*0.3 s*0.2 x2*0.5;
                y with s*0.1;
                y on x2*0.4;
                s on x2*0.1;

                %between c3%
                y*0.4; s*0.2; x3*0.5;
                [y*2.1 s*1 x3*1];
                y with s*0.2;
                y on x3*-0.5;
                s on x3*0.2;
```

Figure 13: Latent v.s. Observed centering comparison: $N_1 = 10$, $N_2 = 10$, $N_3 = 20$

MODEL RESULTS LATENT CENTERING

		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Between LEVEL2 Level								
S	ON							
X		0.100	0.0946	0.0761	0.0697	0.0058	0.910	0.350
Y	ON							
X		0.400	0.3841	0.0823	0.0802	0.0070	0.930	0.990
Between LEVEL3 Level								
S	ON							
X		0.200	0.2395	0.1762	0.2258	0.0323	0.990	0.190
Y	ON							
X		-0.500	-0.4445	0.2514	0.3266	0.0657	0.990	0.260
Intercepts								
S		1.000	0.9484	0.2018	0.2626	0.0430	0.960	0.940

MODEL RESULTS OBSERVED CENTERING

		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Between LEVEL2 Level								
S	ON							
X2		0.100	0.0856	0.0614	0.0643	0.0039	0.970	0.280
Y	ON							
X2		0.400	0.4948	0.0667	0.0725	0.0134	0.720	1.000
Between LEVEL3 Level								
S	ON							
X3		0.200	0.2116	0.1539	0.1918	0.0236	0.980	0.240
Y	ON							
X3		-0.500	-0.3171	0.2267	0.2738	0.0843	0.950	0.240
Intercepts								
S		1.000	0.9757	0.1841	0.2208	0.0342	0.960	0.970

Figure 14: Latent v.s. Observed centering comparison: $N_1 = 20$, $N_2 = 20$, $N_3 = 50$

MODEL RESULTS LATENT CENTERING

		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Between LEVEL2 Level								
S	ON							
X		0.100	0.1052	0.0240	0.0251	0.0006	0.940	0.980
Y	ON							
X		0.400	0.4052	0.0299	0.0297	0.0009	0.930	1.000
Between LEVEL3 Level								
S	ON							
X		0.200	0.1861	0.0962	0.1013	0.0094	0.940	0.460
Y	ON							
X		-0.500	-0.5117	0.1344	0.1468	0.0180	0.980	0.920
Intercepts								
S		1.000	1.0159	0.1143	0.1213	0.0132	0.960	1.000

MODEL RESULTS OBSERVED CENTERING

		Population	ESTIMATES Average	Std. Dev.	S. E. Average	M. S. E.	95% Cover	% Sig Coeff
Between LEVEL2 Level								
S	ON							
X2		0.100	0.0993	0.0221	0.0235	0.0005	0.950	0.970
Y	ON							
X2		0.400	0.4637	0.0278	0.0280	0.0048	0.370	1.000
Between LEVEL3 Level								
S	ON							
X3		0.200	0.1799	0.0898	0.1016	0.0084	0.940	0.450
Y	ON							
X3		-0.500	-0.4647	0.1303	0.1508	0.0181	0.980	0.910
Intercepts								
S		1.000	1.0165	0.1097	0.1237	0.0122	0.970	1.000

6 Real data examples

Here we describe some real data examples and demonstrate what can be learned from using the DSEM3 models. All examples are based on intensive longitudinal data for groups of individuals that would typically be analyzed with two-level DSEM models. The data is analyzed with DSEM3 where an additional level of clustering is included in the model. In the first four examples, the additional level of clustering is the middle level and is based on observations nested within a time period such as observations nested within days. In the fifth examples, the additional level of clustering is the highest level, i.e., individuals are nested within groups/clusters.

The analyses we present here are based on multiple model estimations. Among these are standard 3-level models without autocorrelation as well as DSEM models without the added level of clustering. These model estimations are indispensable preliminary evaluations that help us navigate complex modeling issues. Furthermore, each variable is analyzed separately prior to being added to a multivariate model. Models with random slopes or random autoregressive coefficients are first analyzed with non-random effects. Models with covariates are analyzed with and without the covariates. When covariates are not included we can more easily evaluate the variance decomposition without having to model the variance for the covariates. In many cases, modeling a correlation between two variables is equivalent to modeling a regression between the variables. These alternatives yield equivalent models but often yield different insights. With all of these combinations, each example is based on about a dozen models but we present only one model. That is the DSEM3 model that we found to be of practical interest. Note, however, that not all of the results presented in the text below can be found in the tables but are coming from these preliminary analyses.

In some cases we present results on a standardized scale. Such results are not available in Mplus for DSEM3 models with random slopes or random autoregressive coefficients. This is because for such models the scale of the variables is also random. For DSEM2 models, cluster specific standardization is computed in Mplus and the average standardized results are reported but this is not yet available for DSEM3 models. Without random slopes, DSEM3 model standardization is computed as for standard 3-level models, i.e., the standardization is done on each level separately. However, if we want to see estimated intra-class correlations(ICC) or to decompose the variance in terms of percentages as in (11), the standardization that is needed is based on the

total variance across the three levels. Since such a standardization is not available in Mplus, we standardize the variable prior to model estimation. Such an approach is not optimal since it uses sample variance instead of DSEM3 model estimated variance but nevertheless is quite useful to quickly evaluate the relative importance of the various model components. Such an approach is also available when the model contains random effects.

6.1 Glucose

The data for this example is provided by Jean-Philippe Laurenceau. Glucose levels for 63 individuals are recorded over multiple days, with observations taken every 5 minutes. Because of the short interval, the example is almost continuous-time in nature. If we estimate a DSEM2 model for this example, the autocorrelation is 0.995, i.e., the R^2 of the previous observation is 0.99, which makes the variable difficult to study in terms of other predictors. As the distance between observations becomes very small, the autocorrelation will approach 1. There are two questions that DSEM3 can be used to address in this data. First, what interval can be used instead of the 5-minute interval to make the analysis more meaningful, and second, whether a day effect exists for the glucose data—that is, do glucose levels meaningfully differ from one day to another to support a day-specific effect?

The first DSEM3 analysis that we attempt uses individual as the level 3 clustering variable and hour as the level 2 clustering variable to estimate a simple univariate model as in Section 4.1. This model, however, does not converge, and the autocorrelation estimate converges to 1. If we fix the autocorrelation to 0.9, the model converges, and the residual variance on the within level is estimated to be less than 1% of the variance of the glucose. We thus conclude that the changes that occur within the hour are ignorable, and we can indeed summarize the information using the average hourly glucose value. We can also make the same conclusion using a standard 3-level model (essentially fixing the autocorrelation to 0). In that case, the within-level variance accounts for 4% of the variance, so it is still within ignorable ranges. Nevertheless, both DSEM2 and DSEM3 analyses inform us that the autocorrelation between the values within the hour is very high and the innovation shocks are very small. We conclude that using the average hourly glucose level is a sufficiently good representation of the data.

Next, we attempt to answer the question of whether a day-specific effect exists. Using the average hourly data as the dependent variable, the day vari-

able as the level 2 cluster variable, and the individual as the level 3 cluster variable, we estimate the univariate random autoregressive model. Three covariates are added to the model: weekend indicator, gender indicator, and an age indicator. The Mplus input file is given in Figure 15, and the results are in Figure 16. The results decisively reject the day effect and a weekend effect. Although the weekend effect is positive for both the mean and the autocorrelation, with 63 individuals and just one weekend being observed on average per person, statistical significance is out of reach. The individual-level effect is significant, as both the random intercept and the random autoregressive coefficient yield z-scores of 4 for both variance components. The average autocorrelation is 0.78 for the hourly data. The covariates at the individual level are also not significant at this sample size. The most significant covariate effect is gender on the autocorrelation coefficient, with females showing 2% lower autocorrelation on average, which might be interpretable as females being more reactive to interrupt prolonged periods of too high or too low glucose levels. The z-score for that effect is 1.7. The individual effect accounts for 30% of the variation in the hourly glucose data.

Figure 15: Day effect model for glucose data

```
variable:
names are GMH DAY HOUR GEND AGE WKEND ID HOUR1;
usevar=gmh WKEND GEND AGE;
lagged=gmh(1);
cluster=id day;
between=(day) WKEND;
between=(id) GEND AGE;

data:file=glh.dat;

define:day=day+1;
standardize age;

analysis:  TYPE = threelevel random;
process=2; biter=(10000); estimator=bayes;

model:
%within%
s | gmh on gmh&1;

%between day%
gmh with s;
gmh s on WKEND;

%between id%
gmh with s;
gmh s on gend age;
```

Figure 16: Day effect model for glucose data results

MODEL RESULTS

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%	Significance
Within Level						
Residual Variances						
GMH	0.073	0.001	0.000	0.071	0.075	*
Between DAY Level						
S ON						
WKEND	0.001	0.011	0.472	-0.022	0.021	
GMH ON						
WKEND	0.019	0.026	0.240	-0.030	0.069	
GMH WITH						
S	0.000	0.000	0.391	0.000	0.001	
Residual Variances						
GMH	0.001	0.001	0.000	0.000	0.003	*
S	0.000	0.000	0.000	0.000	0.001	*
Between ID Level						
S ON						
GEND	-0.047	0.028	0.047	-0.102	0.008	
AGE	0.011	0.014	0.207	-0.015	0.038	
GMH ON						
GEND	0.049	0.088	0.283	-0.122	0.222	
AGE	-0.047	0.043	0.133	-0.132	0.037	
GMH WITH						
S	0.023	0.006	0.000	0.013	0.038	*
Intercepts						
GMH	1.562	0.055	0.000	1.457	1.671	*
S	0.796	0.017	0.000	0.761	0.829	*
Residual Variances						
GMH	0.094	0.024	0.000	0.060	0.154	*
S	0.008	0.002	0.000	0.005	0.013	*

6.2 Positive Affect and Tiredness

In this section, we analyze the data described in Muthén, Asparouhov, and Keijsers (2025). Two variables, Positive Affect (PA) and Tiredness, are recorded for 219 individuals at random times, on average about 5 times a day for up to 7 days. We will explore the regression analysis described in Section 4.3, where Tiredness will be used to predict PA. As in the previous example, we include day- and individual-specific effects for both variables, random contemporaneous regression effect, random autocorrelation for both variables, as well as regression for the day specific values and the individual time invariant values. It is of particular interest to see if the three regression coefficients are different across the 3 levels. If these coefficients are different, then we conclude that decomposing the predictor as three separate predictors, one on each of the three levels, improves the model quality compared to the model where the predictor is treated as a whole with just one regression coefficient.

Figure 17 shows the input file for this analysis and Figure 18 shows the results. We did not include the covariances between the various effects to simplify the model. Such a model is typically used as a preliminary analysis to evaluate the need to model the effects as random, however, Asparouhov and Muthén (2024c) show that random effect covariance modeling is essential and if it is ignored may result in biased parameter estimates.

To evaluate the differences across the levels for statistically significance, the difference between the 3 regression coefficients are included in model constraints. These difference are often referred to as the big-fish-little-pond effects; see Marsh et al. (2014). We also standardize the two variables. Because of the random slopes and autocorrelations, model standardization as well as the model-estimated variance/covariance matrix are not available. If the variables are standardized, we can easily see the size of the day effect as a proportion of the unit total variance.

Figure 18 shows that the regression coefficients at all 3 levels are significant and negative as expected. The effect also appears to progress across the levels: the instantaneous effect is smallest, the day effect is larger, and the individual-level effect is the largest. The day effect components are significant, and the z-score for both is above 3. Note that z-scores are not included in the Mplus output with the Bayesian estimator. Statistical significance is instead evaluated with the credibility intervals. However, random effect variance component parameter is an exception to that rule. The credibility

interval for variance parameter will include only positive values since the posterior distribution is obtained from the MCMC generated parameter values which are always positive(proper values). Thus, using the credibility interval for these parameters would always conclude that the parameter is significant. Instead we use the z-score (computed manually from the output) to evaluate statistical significance. We generally use this rule-of-thumb: variance component is significant if the z-score is above 3. This rule-of-thumb uses the higher cutoff value of 3, rather than the usual value of 1.96 because testing at the border of admissible space yields an intractable cutoff value. The additional one unit of standard deviation is used to ensure that we are not at the border of the admissible space. Clearly this rule-of-thumb is not perfect but in our experience it is fairly reliable in terms of yielding the correct conclusion in simulation studies.

The day effect accounts for just 5% of the variance, while the individual-level effect is close to 50% for both variables. One interesting issue to note here is that the regression coefficient on the middle level has a larger standard error than on the highest level, and note that this is counterintuitive. There are 6 times as many observations on the middle level as there are on the highest level, so the order in principle should be reversed. In fact, the corresponding 3-level model (without autocorrelation) does have that kind of ordering in the size of the standard error, i.e., larger at the highest level. The reverse ordering in DSEM3 is in part due to the fact that the day effect is more difficult to identify because it competes with the autoregressive coefficient for explaining covariance between the day observations.

Two of the three differences between the regression coefficients are statistically significant. Only the difference between level 2 and level 3 effects is not significant. It is interesting to note here that if we remove the autoregressive coefficients and estimate the model with a 3-level model, all 3 differences are significant. Ignoring the autocorrelation on the within level results in underestimation of the standard errors as well as incorrect formation of the day-effect size. In the 3-level model, the day effect is overestimated to be above 10% of the variance, double what it should be.

The random effects (regression coefficient and two random autocorrelations) appear to have statistically significant components on both levels. This means that the effect of Tiredness on PA varies across individuals and even across the different days for the same individual. Similarly, the autoregressive coefficient varies across individuals and across days. It is interesting to compare this model also to the model where all the effects are not random. We

find that the day effect, particularly for Tiredness, is much smaller, but not statistically significantly so. To explore this phenomenon, further targeted simulation studies must be conducted where random effects are estimated as fixed. We will not pursue this here, however.

We may also compare the DSEM3 run to the DSEM2 run that ignores the day effect and essentially analyzes all data across the days as one long sequence. Note, however, that four missing data rows are inserted for sleep time, which at the moderate autocorrelation present in this data results in practically independent observations across the different days. Comparisons with DSEM2 (results not included) reveal that even without the day-specific effects, the random slope means remained unchanged, although their variances are larger compared to the variance of DSEM3 on the individual level. Note that this doesn't occur if we compare the models without the random slopes and autocorrelations. In that case, PA autocorrelation in DSEM3 is smaller and there is a larger day effect, while Tiredness autocorrelation doesn't change and the day effect is smaller. We interpret this as evidence that the random effects are needed to properly estimate the day effects and that the heterogeneity of the autocorrelations and the relationship between the two variables is essential for this data.

We also note here that Muthén et al. (2025) analyzed the data with cycles, where the idea is that patterns within-day are repeated across the days. It is possible to include such modeling here as well. One of the advantages of using cycles is that the trigonometric functions, when set up correctly, yield the same patterns across the day because the trigonometric functions are periodic and reset at the beginning of each day. In DSEM3, this sort of repetition of within-day trends occurs naturally. Given the level 2 and level 3 effects, the conditional model for within-day is by definition the same.

Figure 17: Tiredness effect on Positive affect

```
DATA:   FILE = adjusted2.csv;

VARIABLE:
        NAMES = ...
        MISSING = ALL (999);
        USEVAR  = pa tired;
        cluster = id day;
        tinterval = hrs (2 time);
        lagged = pa(1) tired(1);

define: standardize pa tired;

Analysis:
        type = threelevel random;
        estimator = bayes;

Model:
        %within%
        s1 | pa on pa&1;
        s2 | pa on tired;
        s3 | tired on tired&1;

        %between day%
        pa on tired (b2); tired;

        %between id%
        pa on tired (b3); tired;
        [s2] (b1);

Model Constraint:
        new(d1 d2 d3);
        d1=b1-b2;
        d2=b1-b3;
        d3=b2-b3;
```

Figure 18: Tiredness effect on Positive affect

MODEL RESULTS

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%	Significance
Within Level						
Residual Variances						
PA	0.301	0.007	0.000	0.288	0.316	*
TIRED	0.430	0.010	0.000	0.409	0.450	*
Between DAY Level						
PA ON						
TIRED	-0.445	0.153	0.000	-0.860	-0.256	*
Variances						
TIRED	0.056	0.016	0.000	0.028	0.087	*
S1	0.048	0.010	0.000	0.028	0.065	*
S2	0.027	0.008	0.000	0.017	0.044	*
S3	0.037	0.007	0.000	0.026	0.054	*
Residual Variances						
PA	0.033	0.008	0.000	0.019	0.049	*
Between ID Level						
PA ON						
TIRED	-0.547	0.087	0.000	-0.723	-0.383	*
Means						
TIRED	0.084	0.054	0.063	-0.022	0.197	
S1	0.287	0.027	0.000	0.233	0.335	*
S2	-0.181	0.019	0.000	-0.216	-0.144	*
S3	0.415	0.027	0.000	0.362	0.468	*
Intercepts						
PA	0.011	0.051	0.417	-0.085	0.113	
Variances						
TIRED	0.446	0.055	0.000	0.356	0.567	*
S1	0.055	0.013	0.000	0.031	0.084	*
S2	0.027	0.006	0.000	0.016	0.040	*
S3	0.044	0.011	0.000	0.028	0.072	*
Residual Variances						
PA	0.425	0.054	0.000	0.333	0.537	*
New/Additional Parameters						
D1	0.274	0.151	0.002	0.070	0.671	*
D2	0.367	0.090	0.000	0.202	0.546	*
D3	0.084	0.176	0.311	-0.367	0.355	

6.3 Positive Affect Factor Analysis

The positive affect variable described in the previous section is the average score of six 7-category items. In this section we describe a DSEM3 factor analysis where the items are analyzed separately. Detailed description of the measurements is available in Muthén et al. (2025). Three of the items are characterized as low arousal PA and the other three as high arousal PA. EFA analysis reveals that the six item measure two factors: low arousal PA and high arousal PA. One of the items loads about equally well on both factors while the remaining items are pure indicators.

The DSEM3 analysis explores the decomposition of the factors across the three levels: within-day, between-day and between-person levels. The model can be described as follows

$$Y_{pijt} = Y_{1,pijt} + Y_{2,pij} + Y_{3,pj}$$

$$Y_{1,pijt} = \lambda_p \eta_{1,ijt} + \varepsilon_{1,pijt}$$

$$\eta_{1,ijt} = R\eta_{1,ij,t-1} + \xi_{ijt}$$

$$Y_{2,pij} = \lambda_p \eta_{2,ij} + \varepsilon_{2,pij}$$

$$Y_{3,pj} = \nu_p + \lambda_p \eta_{3,j} + \varepsilon_{3,pj}.$$

Equivalently, the model can be written in terms of decomposition of the factors as well as the error terms as follows

$$Y_{pijt} = \nu_p + \lambda_p \eta_{ijt} + \varepsilon_{pijt}$$

$$\eta_{ijt} = \eta_{1,ijt} + \eta_{2,ij} + \eta_{3,j}$$

$$\eta_{1,ijt} = R\eta_{1,ij,t-1} + \xi_{ijt}$$

$$\varepsilon_{pijt} = \varepsilon_{1,pijt} + \varepsilon_{2,pij} + \varepsilon_{3,pj}.$$

In the above equations, η is a vector of size 2, λ_p is a vector of size 2, R is a 2 by 2 autocorrelation matrix for the factors, and ξ_{ijt} is a correlated error terms vector of size 2. The loadings are held equal across the levels so that the factors can be expressed as variables decomposed on the three levels.

Figure 19 contains the Mplus input for estimating the above model. We also include three individual level predictors for the two factors as well as the variable tiredness, which is used as a predictor on all three levels. The decomposition of the factors to a large extend matches the results of the

average PA score found in the previous section. For the low arousal PA, the individual effect accounts for 50% of the variation, and the day effect for 8%. For the high arousal PA, the values are 53% and 7%. These values are slightly higher than those for the observed average PA. This is expected as when item specific errors are removed we expect higher correlations. Cross-lagged effects are not significant and the auto-correlations for the two factors are .29 and .31. The correlations between the two factors across the three levels are .63 for the within-day level, .89 for the between-day level, and .86 for the between-individual level. The between-day correlation is only marginally significantly different from 1, which means that the day specific effect is likely just one dimensional.

The biggest difference between the two factors that we find in this analysis is in the effect of Tiredness on the factors. Tiredness has a much stronger negative effect on the high arousal PA. The correlations between Tiredness and low arousal PA on the three levels are: -.06, -.31, -.37; while for the high arousal PA the values are -.31, -.47, -.53. Among the other predictors the effects on the two factors are similar except for AGE which just like Tiredness has a much stronger negative correlation with the high arousal PA. These results match the findings in Muthén et al. (2025).

Figure 19: Positive affect factor analysis

```
DATA:   FILE = ...;
VARIABLE:  NAMES = ...
           MISSING = ALL (999);
           USEVAR  = PALA1 PALA2 PALA3 PAHA1 PAHA2 PAHA3
           TIRED sex Age SDQ;
           cluster = id day;
           tinterval = hrs (3 time);
           between = (id) sex Age SDQ;

Analysis:
        type = threelevel;
        estimator = bayes;
        biter = (1000);
        proc = 2;

Model:
        %within%
        fw1 by pala1-paha1* (&1 1-4);
        fw2 by paha3* paha1 paha2 (&1 5-7);
        fw1@1;  fw2@1; tired;
        fw1 fw2 on fw1&1 fw2&1 tired;

        %between day%
        fbd1 by pala1-paha1* (1-4);
        fbd2 by paha3* paha1 paha2 (5-7);
        fbd1 fbd2 on tired;

        %between id%
        fbid1 by pala1-paha1* (1-4);
        fbid2 by paha3* paha1 paha2 (5-7);

        fbid1 fbid2 on tired sex Age SDQ;
```

6.4 Smoking Urge and Negative Affect

In this section, we discuss the data described in Muthén, Asparouhov, and Shiffman (2025) and Section 9 of Asparouhov and Muthén (2024a). Smoking Urge (SU) and Negative Affect (NA) are recorded for 235 individuals over a period of 4 weeks, with about 5 observations per day at random times. Previous analysis suggests that the autocorrelation does not carry overnight here as well, and thus using DSEM3 as a modeling framework where the day level is the middle level appears to be suitable. Here we compare DSEM2 and DSEM3 analyses for the VAR model discussed in Section 4.4. We use TINTERVAL with 20-minute periods, which results in nearly 90% missing data inserted to space the observations across periods to match the random times of observations. The results of the two analyses are given in Figures 20 and 21. We see here that on the individual level, the results are largely unaffected by the inclusion of day-specific effects, i.e., DSEM3 and DSEM2 yield the same values. We also see that both variables yield substantial day-specific effects in the DSEM3 analysis (variance parameters z-scores are greater than 10). Since the individual level remains unchanged in the analysis, we conclude that the day effect level interacts primarily with the within-level model. Thus, for the purposes of properly understanding the impact of the day level on the model, we report the ICC of the day level as the proportion of variance not explained by the individual level. The ICC for the SU is 30%, while for NA it is 33%. These kinds of ICC levels are profound, and if ignored, will have a substantial effect on the model. The DSEM2 autocorrelation estimates of 0.75 and 0.46 are estimated in DSEM3 to be much lower: 0.57 and 0.09. The cross-lagged effect from NU to SU, on the other hand, increase from 0.18 to 0.28.

An alternative model where the three correlations in DSEM3 are replaced with the regression of smoking urge on negative affect is also estimated. This model also includes a contemporaneous relationship between the variables rather than just cross-lagged. The standardized results for this model are presented in Figure 22. We see that both cross-lagged relationships are not significant. In contrast, for DSEM2, the cross-lagged relationships are significant with and without the contemporaneous effect. Since the DSEM2 model is nested within the DSEM3 model, the results of DSEM3 are presumed more reliable and we conclude that the cross-lagged relations are not needed and that the contemporaneous relationship on the three levels is a sufficient representation for this data. It is interesting to note here that the standardized

results reveal that the relationship between the variables strengthens across the levels: it is the strongest on the individual level, second strongest on the day level, and weakest on the within-day level.

Next we want to clarify why DSEM2 is nested within DSEM3. Based on an average sleep of 8 hours and Tinterval of 20 minutes, the DSEM2 analysis has at least 24 rows of missing data inserted between the last observations in one day and the first observation of the next days. The correlation between these observations in DSEM2 is R^{24}, where R is the autocorrelation matrix given in Figure 21. That matrix is practically zero with the highest entry in it being 0.001. Without using matrix computations, one can also make the same conclusion by considering the higher of the two autocorrelation given in Figure 21. That is the autocorrelation for Negative Affect: 0.749. Simply using a calculator, we compute that $0.749^24 \approx 0.001$. Autocorrelations less than 0.01 can be considered practically zero. Autocorrelations of 0.01 implies an R^2 of the prior observation of 0.0001. In summary we see that in the DSEM2 model, the observations across days are practically independent. In such situations, the difference between the DSEM3 model and the DSME2 model is only in the day effects. Therefore we can make the claim that DSEM3 is nested above DSEM2. Note, however, that such a statement is not true in general, and if we want to make that statement we must establish as we did above implied independence of the observations across the different days.

Figure 20: Smoking Urge and Negative Affect DSEM3

MODEL RESULTS

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%	Significance
Within Level						
URGE ON						
URGE&1	0.088	0.012	0.000	0.063	0.110	*
NEGAFF&1	0.277	0.029	0.000	0.212	0.320	*
NEGAFF ON						
URGE&1	0.007	0.005	0.047	-0.001	0.018	
NEGAFF&1	0.565	0.009	0.000	0.550	0.585	*
URGE WITH						
NEGAFF	0.199	0.010	0.000	0.178	0.216	*
Residual Variances						
URGE	2.946	0.034	0.000	2.887	3.033	*
NEGAFF	0.328	0.005	0.000	0.316	0.336	*
Between DAY Level						
URGE WITH						
NEGAFF	0.125	0.012	0.000	0.101	0.147	*
Variances						
URGE	1.125	0.039	0.000	1.045	1.201	*
NEGAFF	0.208	0.007	0.000	0.195	0.221	*
Between SUBJECT Level						
URGE WITH						
NEGAFF	0.611	0.110	0.000	0.410	0.855	*
Means						
URGE	3.660	0.138	0.000	3.382	3.934	*
NEGAFF	0.049	0.042	0.093	-0.028	0.141	
Variances						
URGE	5.010	0.508	0.000	4.181	6.076	*
NEGAFF	0.393	0.040	0.000	0.323	0.471	*

Figure 21: Smoking Urge and Negative Affect DSEM2

MODEL RESULTS

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%	Significance
Within Level						
URGE ON						
URGE&1	0.455	0.028	0.000	0.391	0.482	*
NEGAFF&1	0.183	0.020	0.000	0.151	0.227	*
NEGAFF ON						
URGE&1	0.007	0.002	0.000	0.003	0.011	*
NEGAFF&1	0.749	0.008	0.000	0.725	0.756	*
URGE WITH						
NEGAFF	0.159	0.008	0.000	0.151	0.183	*
Residual Variances						
URGE	3.201	0.096	0.000	3.086	3.415	*
NEGAFF	0.291	0.007	0.000	0.285	0.311	*
Between Level						
URGE WITH						
NEGAFF	0.606	0.098	0.000	0.410	0.785	*
Means						
URGE	3.651	0.149	0.000	3.397	3.952	*
NEGAFF	0.054	0.042	0.103	-0.026	0.135	
Variances						
URGE	5.119	0.499	0.000	4.231	6.039	*
NEGAFF	0.400	0.038	0.000	0.331	0.480	*

Figure 22: Contemporaneous regression of Smoking Urge and Negative Affect DSEM3

STANDARDIZED MODEL RESULTS

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%	Sig
Within Level						
URGE ON						
URGE&1	0.080	0.018	0.000	0.031	0.104	*
NEGAFF&1	-0.019	0.014	0.066	-0.044	0.006	
NEGAFF	0.238	0.009	0.000	0.220	0.255	*
NEGAFF ON						
URGE&1	0.015	0.014	0.130	-0.013	0.043	
NEGAFF&1	0.565	0.009	0.000	0.546	0.579	*
Residual Variances						
URGE	0.938	0.004	0.000	0.928	0.945	*
NEGAFF	0.676	0.009	0.000	0.660	0.694	*
Between DAY Level						
URGE ON						
NEGAFF	0.260	0.022	0.000	0.213	0.300	*
Variances						
NEGAFF	1.000	0.000	0.000	1.000	1.000	
Residual Variances						
URGE	0.933	0.011	0.000	0.910	0.954	*
Between SUBJECT Level						
URGE ON						
NEGAFF	0.425	0.056	0.000	0.311	0.528	*
Variances						
NEGAFF	1.000	0.000	0.000	1.000	1.000	
Residual Variances						
URGE	0.820	0.047	0.000	0.721	0.903	*

6.5 Emotional Cost and Interest for University Students

The following data was provided by Patrick Beymer and consists of weekly surveys of 5,407 university students enrolled in 62 classes in a total of 13 science courses. The data is collected for up to 13 weeks. The two variables that we consider here are emotional cost of course load and interest in the course. Six individual-level covariates are available: final course grade, high school GPA, gender, race, first-generation college status, and self-declared STEM intention. We would like to conduct a two-level DSEM, i.e. DSEM3, where we account for the nesting of the students in sections/classrooms or courses. The purpose can be two-fold. One can consider the course/classroom-specific effect to be a variable of interest. For both of these variables, evaluating the average level of interest being generated in each classroom as well as the average emotional cost to the students can be of interest for many reasons. On the other hand, if we want to consider the effect of high school GPA on emotional cost to the student, it would clearly be a deficiency to ignore the class and course information. Not only would the standard errors be underestimated by ignoring the extra level of clustering, but also there is potential for the point estimates to be subject to confounding effects such as "high school GPA" might be associated with "heavier emotional cost classes." This example is different from the examples considered previously. Here, the individual is the middle-level clustering variable, while students will be nested in course or classroom. As in any multilevel analysis, there are two interpretations of the model. The standard interpretation is that the higher level of clustering provides an effect that makes the variables within the cluster more correlated. The second interpretation is that we account for non-invariance across the various groups of individuals.

The first question we want to consider is whether the course nesting or classroom nesting is the correct higher-level nesting variable. Is the student outcome of interest driven primarily by the course variable or by the teacher/classroom variable. If the course variable doesn't have a substantial impact, or a statistically significant impact, or it has an impact that is smaller than the classroom variable, we clearly should be using the classroom variable as the higher-level nesting variable. To investigate this issue we conduct preliminary analysis similar to Section 4.2—a univariate DSEM3 model with random autocorrelation using both variables and both nesting options. The outcome is that course effects are not significant (z-score of the variance

parameter of the random intercept at the course level is less than 3), while classroom effects are significant. The size of the effects, however, is similar. Due to the fact that there are only 13 courses in this data, while classroom entries are 62, this is likely just a sample size effect. Because the number of classrooms is larger, the evidence for classroom effect is more solid than that for the course variable. We continue the analysis using the classroom as the higher-level clustering variable.

Next, we want to illustrate the effect of accounting for the classroom effect on the individual-level regression. We will use the emotional cost variable for this illustration. From the univariate analysis, we see that the classroom level autocorrelation effect is not significant. Thus, we remove that effect, and the random autocorrelation is estimated as an individual-level variable only. We estimate the effect of the 6 individual-level covariates on the emotional cost variable. The DSEM3 input is given in Figure 23, and the comparison of DSEM2 and DSEM3 results is given in Figure 24. A substantially different pattern of significance emerges. In DSEM2, 4 of the covariates have significant effects on the autocorrelation, although the Z-score of all of these is less than 3 and thus can be considered marginal. These effects are all considered not significant in DSEM3. DSEM3 concludes that emotional cost is higher for non-white and first-generation college students. These effects were not significant in DSEM2. The opposite occurred for STEM intent. Also notable is the difference in the size of the effect of high school GPA. This effect is substantially higher in the DSEM3 model. We see here that accounting for the classroom effect shifts the predictive power of the covariates.

We also consider the VAR model for the two variables: course interest and emotional cost. The standardized DSEM3 results are given in Figure 25. We see that the two variables are negatively correlated on all the levels, and the correlation increases with the levels: within-student weekly correlation is -0.09, average individual values correlation is -0.43, and the average class-level correlation is -0.75. DSEM3 cross-lagged results are identical to DSEM2 cross-lagged results: small negative but significant in both directions. The total autocorrelation from the values in the previous period accounts for about 7% of the variation within-student. The classroom-level effects for emotional cost is 13% of the total variance, and for the course interest variable it is 9%. As usual, these values are computed from the decomposition given in (11).

Figure 23: Emotional cost of university classes DSEM3

```
DATA: file=reduced4.dat;

VARIABLE:
names = Course id week section  female  fg race
tecost  oecost  lvcost  emcost  interst
stemint fggpa   hsgpa;
cluster = section id;
USEVAR = emcost stemint fggpa   hsgpa   female  fg race;
BETWEEN =  (id) stemint fggpa   hsgpa   female  fg race;
MISSING = all(-999);
LAGGED = emcost(1);
TINTERVAL = week(1);

ANALYSIS:
TYPE IS threelevel random;
estimator = bayes; processors = 2;

MODEL:

%WITHIN%
phi | emcost on emcost&1;

%BETWEEN ID%
emcost with psi; [phi];
emcost phi on stemint fggpa     hsgpa   female  fg race;
stemint fggpa   hsgpa   female  fg race;

%BETWEEN section%
emcost; [phi@0]; phi@0;
```

Figure 24: Emotional cost of university classes DSEM3 vs. DSEM2

MODEL RESULTS

		Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%	Sig
Between ID Level DSEM3							
PHI	ON						
STEMINT		0.011	0.008	0.072	-0.004	0.027	
FGGPA		-0.015	0.011	0.093	-0.037	0.010	
HSGPA		0.003	0.034	0.452	-0.063	0.072	
FEMALE		-0.003	0.016	0.413	-0.038	0.027	
FG		0.016	0.022	0.225	-0.028	0.059	
RACE		-0.047	0.023	0.028	-0.089	0.001	
EMCOST	ON						
STEMINT		0.027	0.017	0.068	-0.008	0.057	
FGGPA		-0.486	0.026	0.000	-0.533	-0.430	*
HSGPA		1.097	0.080	0.000	0.901	1.206	*
FEMALE		0.271	0.041	0.000	0.186	0.343	*
FG		0.185	0.055	0.000	0.079	0.290	*
RACE		0.191	0.055	0.000	0.072	0.291	*
Between Level DSEM2							
PHI	ON						
STEMINT		0.013	0.007	0.007	0.001	0.026	*
FGGPA		-0.022	0.011	0.017	-0.046	-0.004	*
HSGPA		0.066	0.027	0.003	0.019	0.120	*
FEMALE		0.002	0.017	0.427	-0.037	0.036	
FG		0.018	0.023	0.197	-0.020	0.067	
RACE		-0.059	0.025	0.003	-0.110	-0.012	*
EMCOST	ON						
STEMINT		-0.096	0.016	0.000	-0.131	-0.067	*
FGGPA		-0.448	0.023	0.000	-0.492	-0.406	*
HSGPA		0.126	0.058	0.013	0.020	0.241	*
FEMALE		0.338	0.037	0.000	0.265	0.411	*
FG		0.055	0.056	0.147	-0.056	0.165	
RACE		0.007	0.050	0.423	-0.087	0.111	

Figure 25: DSEM3 VAR for Emotional cost and Interest standardized results

STANDARDIZED MODEL RESULTS DSEM3 VAR

	Estimate	Posterior S.D.	One-Tailed P-Value	95% C.I. Lower 2.5%	Upper 2.5%	Sig
Within Level						
EMCOST ON						
EMCOST&1	0.239	0.005	0.000	0.228	0.248	*
INTERST&1	-0.073	0.005	0.000	-0.085	-0.063	*
INTERST ON						
EMCOST&1	-0.061	0.006	0.000	-0.071	-0.049	*
INTERST&1	0.241	0.005	0.000	0.231	0.251	*
INTERST WITH						
EMCOST	-0.089	0.005	0.000	-0.099	-0.079	*
Residual Variances						
EMCOST	0.933	0.003	0.000	0.928	0.938	*
INTERST	0.935	0.003	0.000	0.929	0.939	*
Between ID Level						
EMCOST WITH						
INTERST	-0.429	0.012	0.000	-0.454	-0.409	*
Variances						
EMCOST	1.000	0.000	0.000	1.000	1.000	
INTERST	1.000	0.000	0.000	1.000	1.000	
Between SECTION Level						
EMCOST WITH						
INTERST	-0.729	0.092	0.000	-0.865	-0.489	*
Means						
EMCOST	6.316	0.715	0.000	5.172	8.011	*
INTERST	9.735	1.082	0.000	7.721	11.926	*
Variances						
EMCOST	1.000	0.000	0.000	1.000	1.000	
INTERST	1.000	0.000	0.000	1.000	1.000	

7 Conclusion

Dynamic structural equation modeling is becoming more popular. More intensive data is being collected and analyzed. As the data frame increases, more data features must be addressed in the model. The DSEM3 modeling framework expands on that front and adds one more level of nesting. In this article, we made an attempt to motivate the use of this expanded framework and suggest a variety of research questions that could be pursued. The new models can be compared with the simpler 3-level SEM models and the DSEM models to more fully understand the data. Four correlation modeling techniques: autoregressive, level 1 effects, level 2 effects, and level 3 effects; compete and interact with each other in the DSEM3 framework. Simulation studies can be used to further illuminate various aspects of the modeling.

References

[1] Asparouhov, T., & Muthén, B. (2006). Constructing covariates in multilevel regression. Mplus Web Notes, 11, 1–8.

[2] Asparouhov, T., Hamaker, E.L. & Muthén, B. (2018). Dynamic structural equation models. Structural Equation Modeling: A Multidisciplinary Journal, 25, 359-388.

[3] Asparouhov, T., & Muthén, B., (2019). Latent variable centering of predictors and mediators in multilevel and timeseries models. Structural Equation Modeling: A Multidisciplinary Journal. 26, 119–42.

[4] Asparouhov, T. & Muthén, B. (2024a). Continuous Time Dynamic Structural Equation Models.
https://www.statmodel.com/download/CTRDSEM.pdf

[5] Asparouhov, T., & Muthén, B., (2024b). Practical Aspects of Dynamic Structural Equation Models. https://www.statmodel.com/download/PDSEM.pdf

[6] Asparouhov, T., & Muthén, B., (2024c). Covariance of Random Effects in Multilevel Modeling. Web Note No. 24. https://www.statmodel.com/examples/webnotes/Webnote%2024.pdf

[7] Hamaker, E. L., Asparouhov, T., and Muthén, B. (2023). Dynamic structural equation modeling as a combination of time series modeling, multilevel modeling, and structural equation modeling. In R. H. Hoyle (Ed.), The handbook of structural equation modeling (2nd ed., pp. 576–596). Guilford.

[8] Hamaker E.L. and Grasman R.P.P.P. (2015). To center or not to center? Investigating inertia with a multilevel autoregressive model. Frontiers in Psychology., 5, 1492.

[9] Lüdtke, 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, 203-229.

[10] Marsh, H.W., Kuyper H., Morin A., Parker P., Seaton M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. Learning and Instruction, 33, 50-66

[11] Muthén B., T. Asparouhov, and L. Keijsers (2025). Dynamic Structural Equation Modeling With Cycles." Structural Equation Modeling: A Multidisciplinary Journal 32, 264–286.

[12] Muthén B., Asparouhov, T., and Shiffman, S. (2025). Dynamic structural equation modeling with floor effects. Psychological Methods. https://dx.doi.org/10.1037/met0000720

[13] Muthén B. & Satorra, A (1995). Complex sample data in structural equation modeling. Sociological methodology, 25, 267-316.

[14] Nickell, S. (1981). Biases in dynamic models with fixed effects. Econometrica: Journal of the Econometric Society, 1417-1426.