Practical Aspects of Dynamic Structural Equation Models

Tihomir Asparouhov and Bengt Muthén

Mplus

November 17, 2021

1 Introduction

In this note we address some practical aspects of dynamic structural equation models. In particular, we focus on what can be done when the model estimation fails, i.e., it didn't converge. Here we provide a check-list of ideas that may be helpful in resolving the convergence problems and may help with further understanding the analysis.

2 The source of the problem

Generally, there are two types of common error messages produced by Mplus in failed DSEM analysis. The first one is that the convergence criterion was not satisfied within the maximum number of iterations. By default the maximum number of iterations is 50000. It is possible to increase that number with the BITER option, however, that generally would be the last resort attempt to resolve the issue. When this problem occurs, the estimation is mixing very slowly. The proper way to fix the problem is to identify the causes for the slow mixing and to modify the model or data accordingly. The second common error message is that the between level variance covariance matrix is singular or that it could not be updated due to a singular posterior distribution. This situation could be more complex to diagnose but generally the issue is most likely related to the random effects in the model.

It many situations, the source of the problem is not easily identifiable and there are several causes combined leading to convergence problems. Below we list some common scenarios that occur in real data applications.

2.1 Unidentified or poorly identified model

If the model is unidentified or is poorly identified, the outcome of the estimation very likely will be non-convergence within the maximum number of iterations. For example, latent variables measured by one or two indicators are often poorly identified and may be the source of the problem. The solution in this case is to use a better identified model by introducing model parameter restrictions, such as loadings fixed to 1 in the case of the two-indicator factor model.

2.2 Random effects with small variances

If the variance of the random effect is small, the variance of the conditional distribution used for updating the random effect will likely be small as well. That means that the random effect might not change much across the MCMC iterations. This in turn will cause the parameters related to that random effect, such as the mean and the regression parameters of that random effect on other variables, to be mixing very slowly and possibly result in non-convergence. Random effects with small variances can be converted to non-random effects. In some situations, a small variance random effect can remain in the model if we simply do not correlate the random effect with other random effects. By default, Mplus will set all random effects on the between level to be uncorrelated with each other. When the random effect are uncorrelated, the model is more likely to converge. In addition, the model with uncorrelated random effects can be used to identify those small variance random effects that are not needed in the model, before proceeding with a more advanced model that correlates all random effects. Small variance random effects tend to cause convergence problems when cluster sizes are larger and when the random effects are correlated with other variables.

We illustrate the small variances convergence problems with a simulation study. Figure 1 contains the Mplus montecarlo setup for a bivariate two-level DSEM model. The data generating DSEM model has random intercepts and non-random auto-regressive (AR) coefficients, which include the cross-lagged coefficients. The estimated DSEM model has random intercepts and random AR coefficients. The variances of the AR random effects are all zero because they are not random in the data generating model. We can expect that the estimated values will also be near zero and therefore the simulation study illustrates small variance random effects. In the estimated between level model in Figure 1, all random effects are correlated. In this study, 3 out of 5 replications did not converge. If we replace the fully correlated between level model with the model where only the random intercepts are correlated and the small variance AR random effects are uncorrelated among each other and the random intercepts, the estimation converges in all 5 replications. Such a result shows that small variance random effects cause problems mainly when they are correlated with other variables.

Figure 1: Simulation study illustrating small variances convergence problems

```
MONTECARLO:  NAMES ARE y z;
NOBS = 720;  NREP = 5;
NCSIZES = 2; CSIZES = 30(20) 20(6);
lagged = y(1) z(1);
missing=all;

model missing: [y*0 z*0];

ANALYSIS:   TYPE = TWOLEVEL RANDOM;
estimator=bayes;  proc=2; biter=(30000);

MODEL POPULATION:
  %WITHIN%
  y on y&1*0.7;
  z on z&1*0.7;
  z on y&1*0.2;
  y on z&1*0.2;
  y*0.5 z*0.5;
  y with z*-0.25;

  %BETWEEN%
  y*1; z*1;
  y with z*0.8;

MODEL:
  %WITHIN%
  s1 | y on y&1;
  s2 | z on z&1;
  s3 | z on y&1;
  s4 | y on z&1;

  %BETWEEN%
  y z s1 s2 s3 s4 with y z s1 s2 s3 s4;
```

2.3 Non-stationarity

A model is stationary when the eigenvalues of the AR (auto-regressive) matrix are all less than 1. In the univariate case this translates to the auto-regressive coefficient being less than 1. In the multivariate case, it is not easy to determine if the AR matrix has all eigenvalues less than 1 simply because the eigenvalues are not easy to compute by hand. In the bivariate case, however, we can provide a little guidance to quickly determine if the AR matrix is stationary. Suppose that the AR matrix is

$$R = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}.$$

The auto-regressive coefficients are r_{11} and r_{22} and the cross-lagged coefficients are r_{12} and r_{21}. In the most common scenario, where the AR coefficients r_{11} and r_{22} are positive and less than 1, and the cross-lagged coefficients are smaller than the auto-regressive coefficients, $r_{12} < r_{11}$ and $r_{21} < r_{22}$, we can simply check that the sum of the coefficients in each equation/row of the AR matrix are less than 1. That is, we can show that if $r_{11} + r_{12} < 1$ and $r_{22} + r_{21} < 1$, the model is stationary. The eigenvalues x_1 and x_2 satisfy

$$x_1 + x_2 = r_{11} + r_{22}$$

$$x_1 x_2 = r_{11} r_{22} - r_{12} r_{21}$$

Under the above restrictions of the autoregressive parameters, we can see that $r_{12} r_{21} < (1 - r_{11})(1 - r_{22})$. Because of the above equations, the last inequality is equivalent to $(1 - x_1)(1 - x_2) > 0$, which in turn leads to the conclusion that both eigenvalues must be between 0 and 1 and that the model is stationary.

In principle, stationarity is not required in DSEM. However, when stationarity is not satisfied in the entirety of the posterior distribution of the auto-regressive matrix, convergence problems can occur. In the univariate case, the MCMC updating involves dividing by $1-r$. If r is close to 1, numerically this is division by zero and the computation will be contaminated with large numerical error. During the MCMC iterations, the entire posterior distribution must be visited. If the posterior distribution is wide enough to cross over into non-stationarity, given a sufficient number of MCMC iterations, the posterior draw of r will eventually be close enough to 1 and destabilize the computation. This exact thing can happen in the multivariate case. The

matrix I minus the AR matrix would become singular and the posterior distribution of the intercept would be undefined. Furthermore, when the AR matrix is cluster specific, the AR matrix must remain stationary in every cluster in the entire posterior distribution, i.e., in every MCMC draw. When a particular cluster has very limited information, due to a poorly identified model or a limited number of observed values, the posterior distribution of the AR matrix will be quite wide and quite likely to lead to non-stationary values. This in turn may cause convergence problems.

The one key feature that appears to consistently separate montecarlo studies and real data applications is the size of the variances of the random auto-regressive coefficients. In montecarlo studies, the variances of the random AR effects are usually set between 0.01 and 0.1. Higher variances typically will lead to non-stationarity, at least in some of the clusters. In real data applications, it is fairly common, however, to see the variances of the random AR effects above 0.1. Typically, this would imply that the range of the AR coefficient across clusters is of length at least 0.6 and that most likely, the point estimates in some of the clusters are not stationary. Note that this is an observation regarding just the point estimates. The range of the entire posterior distribution would be even larger. In such situations, the conclusion is that, it is a matter of getting through enough MCMC iterations and chains to run into a set of parameters that will destabilize the computation through dividing by a near-zero number. To summarize, variances of random AR effects above 0.1 should be inspected carefully in each cluster.

We do not view the issues discussed above as stationarity problems. In these situations, we would not recommend modeling trends or striving for stationary models. We view the problem instead as a limited information problem. Weak information leading to wide posterior distributions for a few of the clusters that can cause computational problems for the entire estimation.

We illustrate the non-stationarity convergence problems with a simulation study. Figure 2 contains the Mplus montecarlo setup for a bivariate two-level DSEM model with random intercept and random AR matrix. With the average random effect parameters, the model is stationary but close to the stationarity borderline. In this study, 5 out of 10 replications did not converge. For the 5 iterations that converged, the average random intercept variance estimates are very large and don't match the true values at all. The means of the diagonal AR random effects in Figure 2 are set at 0.7. If we change these means to 0.4 in the data generation model, the model is moved

away from the stationarity border. In that case, the estimation converges in 9 out of the 10 replications and the parameter estimates are close to the true values.

Figure 2: Simulation study illustrating non-stationarity convergence problems

```
MONTECARLO:  NAMES ARE y z;
NOBS = 1000; NREP = 10;
NCSIZES = 1; CSIZES = 100(10);
lagged = y(1) z(1);
missing=all;

model missing: [y*0 z*0];

ANALYSIS:   TYPE = TWOLEVEL RANDOM;
estimator=bayes; proc=4; biter=(30000); chain=4;

MODEL POPULATION:
  %WITHIN%
  s1 | y on y&1;
  s2 | z on z&1;
  s3 | z on y&1;
  s4 | y on z&1;
  y*0.5 z*0.5; y with z*0.25;

  %BETWEEN%
  y*1; z*1; y with z*0.8;
  s1-s4*0.01; [s1-s2*0.7 s3-s4*0.25];

MODEL:
  %WITHIN%
  s1 | y on y&1;
  s2 | z on z&1;
  s3 | z on y&1;
  s4 | y on z&1;
  y*0.5 z*0.5; y with z*0.25;
```



```
  %BETWEEN%
  y*1; z*1; y with z*0.8;
  s1-s4*0.01; [s1-s2*0.7 s3-s4*0.25];
```

2.4 Large number of random effects compared to the size of the clusters

Large number of random effects compared to the size of the clusters typically will produce wide posterior distribution. Such large posterior distributions may lead to a non-stationary AR matrix that can cause convergence problems. The most reasonable approach in this case is to reduce the number of random effects. Ideally, the number of random effects should be smaller than the size of the clusters. If the auto-regressive values are small, however, it is possible to estimate models with larger number of random effects.

2.5 Large amount of missing data

Large amount of missing data combined with small amount of observed data can also lead to large posterior distributions and non-stationary AR matrix. In this case as well, the most reasonable solution is to convert some of the cluster-specific random effects to cluster-invariant non-random effects. If only few clusters have excessive amount of missing data, such clusters can be removed from the sample. These clusters contain very small amount of information and removing them is unlikely to affect the estimation in a meaningful way. Simulation studies show that removing clusters with very limited information may increase MSE for some parameters but will also decrease MSE for other parameters.

2.6 Clusters that have very few observed values

Clusters that have very few observed values, i.e., short time-series data, will naturally yield wide posterior distribution for the random effects. This in turn may yield non-stationary convergence problems. The preferred solution in this case is to remove small clusters from the sample. As in the previous case, the overall performance of the model estimation will not change dramatically but it can resolve convergence problems.

We illustrate the small clusters convergence problems with a simulation study. Figure 3 contains the Mplus montecarlo setup for a bivariate two-level DSEM model with random intercept, random AR matrix, and random variance covariance matrix. 20 small clusters are included in this montecarlo study. In this study, 3 out of 5 replications did not converge. If we remove

the small clusters from the simulation study, the estimation converges in all replications.

Figure 3: Simulation study illustrating small clusters convergence problems

```
MONTECARLO:  NAMES ARE y z;
NOBS = 1040; NREP = 5;
NCSIZES = 2; CSIZES = 50(20) 20(2);
lagged = y(1) z(1);
missing=all;

model missing: [y*0 z*0];

ANALYSIS:   TYPE = TWOLEVEL RANDOM;
estimator=bayes; proc=4; chain=4; biter=(30000);

MODEL POPULATION:
  %WITHIN%
  s1 | y on y&1;
  s2 | z on z&1;
  s3 | z on y&1;
  s4 | y on z&1;
  v1 | y; v2 | z;
  f by y@1 z@1;
  v3 | f;

  %BETWEEN%
  y*1; z*1; y with z*0.8;
  s1-s4*0.005; v1-v3*0.005;
  [v1-v3*-1]; [s1-s2*0.4 s3-s4*0.2];

MODEL:
  %WITHIN%
  s1 | y on y&1;
  s2 | z on z&1;
  s3 | z on y&1;
  s4 | y on z&1;
  v1 | y; v2 | z;
  f by y@1 z@1;
  v3 | f;

  %BETWEEN%
  y z s1-s4 v1-v3 with y z s1-s4 v1-v3;
  [v1-v3*-1]; [s1-s2*0.4 s3-s4*0.2];
```

2.7 Random residual covariance with varying signs

Hamaker et al. (2018) illustrates how random covariance can be estimated in Mplus by using the random variance for a common factor. This approach, however, is mostly applicable when the random covariances have the same sign across the clusters. If the covariances are positive in some clusters and negative in other clusters, the construction may fail. One way to resolve the problem is to remove the random covariance from the model.

The sign of the covariance is determined by the loadings of the common factor. If both loadings are set to 1, the covariance will be positive. If one loadings are set to 1 and the other one to -1, the covariance will be negative.

There are three separate variations that can occur.

- The residual covariance is not random but is modeled as random. Here one can expect that the random effect, log of the residual variance, will have zero variance on the between level causing singular between level matrix.

- The residual covariance is random but it takes positive and negative values across the clusters. Since the covariance is modeled as either positive or negative that means that the covariance will be converging to 0 for those that are of the sign not accommodated by the model. That in turn will results in the random effect on the between level being minus infinity, causing instability in the estimation.

- The residual covariance is random and positive but is modeled as negative, or the residual covariance is random and negative but is modeled as positive. Since the random parameter is residual covariance, it is difficult to know in advance which sign is most appropriate. In this case of course there is an easy solution. The model should be switched to match the correct sign. In practical settings, both models can be estimated for comparison.

We illustrate the random residual covariance with varying signs convergence problems with a simulation study. Figure 4 contains the Mplus montecarlo setup for a two-level model with random covariance. In the data generation model, the random covariance is generated as the difference between the random variances of the two factors f_1 and f_2. The random covariance is equally likely to be positive or negative. The estimated model assumes that

the random covariance is positive. In this study, 5 out of 10 replications did not converge. If we remove the factor f_2 from the data generation, the model estimation converges in all 10 replications.

Figure 4: Simulation study illustrating the random residual covariance with varying signs convergence problems

```
MONTECARLO:  NAMES ARE y z;
             NOBS = 1000;
             NREP = 10;
             NCSIZES = 1;
             CSIZES = 50(20);

ANALYSIS:   TYPE = TWOLEVEL RANDOM;
                estimator=bayes;
            proc=2;
            biter=(30000);

MODEL POPULATION:

  %WITHIN%
  y*1 z*1;
  f1 by y@1 z@1;
  f2 by y@1 z@-1;
  v1 | f1; v2 | f2;

  %BETWEEN%
  y*1; z*1;
  y with z*0.8;
  v1*0.1 v2*0.1;

MODEL:

  %WITHIN%
  v1 | y; v2 | z;
  f by y@1 z@1;
  v3 | f;

  %BETWEEN%
```

2.8 Exploding parameter estimates

In some situations, convergence failure results in final estimates that show exploding parameter values, that is, very large parameter values. This can occur when the model is unidentified or poorly identified. The exploding parameters could be analyzed to identify the weakness of the model. Exploding parameters can also occur when there is a singular variance covariance matrix. In that case, the instability can spread in the entire model and analyzing the exploding parameters is unlikely to provide useful information.

2.9 Variance fixed to zero or to a small value

In some modeling situations it is necessary to fix a residual variance to zero or a small value. This may cause very slow mixing in the Bayes estimation. The problem can be resolved by increasing the small value the variance is fixed to.

2.10 The random walk option

The Mplus default analysis options usually yield the most efficient model estimation. If estimation options are specified that are different from the Mplus defaults, these options might be the source of the problem. One example is the option ALGO=GIBBS(RW). This option allows Mplus to estimate non-block diagonal variance covariance matrices, however, it can produce very slow mixing estimation. We recommenced using models with block diagonal variance covariance matrices and removing that option. The random walk option works well only when the size of the variance covariance matrix is small.

Consider the bivariate DSEM simulation study given in Figure 5. In this simulation study, the intercepts, the auto-regressive parameters, and the residual variance covariance matrices are all random effects. A between level covariate X is also included in the model. The covariate is used to predict the dependent variable Y, using the Y on X statement. The estimated DSEM model also includes all covariances between the random effects as well as all the covariances between the random effects and the covariate X, with the exception of the covariance between Y and X which would be unidentified given the regression model between the two variables. Because the estimated variance covariance matrix on the between level includes all

the parameters except one, it is not a block diagonal matrix and the option ALGO=GIBBS(RW) must be specified. Figure 6 shows the tech1 output for this model and illustrates why the variance covariance matrix is not block-diagonal. The covariance between X and Y is fixed to 0 and this violates the block-diagonal structure. A variance covariance matrix is block-diagonal if it satisfies the following property: if A is correlated with B and C then B and C must also be correlated. In this model, this property is not satisfied for the variables X, Y and Z.

In this simulation study given in Figure 5, all 5 replications fail to converge. The solution to this problem is to replace the covariances **s1-s4 v1-v3 z with x;** with the regressions **s1-s4 v1-v3 z on x;**. Such a model has a block-diagonal variance covariance matrix and the option ALGO=GIBBS(RW) can be removed. All 5 replications converge with this modification. The two DSEM models are equivalent but only the second model can be estimated efficiently.

Figure 5: Simulation study illustrating the random walk/GIBBS(RW) convergence problems

```
MONTECARLO:  NAMES ARE y z x;
NOBS = 1000; NREP = 5;
NCSIZES = 1; CSIZES = 50(20);
lagged = y(1) z(1);
between=x;

ANALYSIS:   TYPE = TWOLEVEL random;
estimator=bayes; proc=2; ALGORITHM=GIBBS(RW); biter=5000;

MODEL POPULATION:
  %WITHIN%
  s1 | y on y&1;
  s2 | z on z&1;
  s3 | z on y&1;
  s4 | y on z&1;
  f by y@1 z@1;
  v1 | y; v2 | z; v3 | f;

  %BETWEEN%
  y*1 z*1 x*1 s1-s4*0.01 v1-v3*0.01;
  y on x*0.4; z with y*0.3; z with x*0.2;
  [v1-v3*-1 s1-s2*0.4 s3-s4*0.2];

MODEL:

  %WITHIN%
  s1 | y on y&1;
  s2 | z on z&1;
  s3 | z on y&1;
  s4 | y on z&1;
  f by y@1 z@1;
  v1 | y; v2 | z; v3 | f;

  %BETWEEN%
  y on x;
  s1-s4 v1-v3 y z with s1-s4 v1-v3 y z;
  s1-s4 v1-v3 z with x;
```

Figure 6: Tech1 with non-block diagonal variance covariance

PSI

	V2	V3	X	Y	Z
V2	32				
V3	38	39			
X	45	46	47		
Y	53	54	0	55	
Z	61	62	63	64	65

2.11 Excessive number of iterations

The posterior distribution of the random effects and the model parameters is generally concentrated around their point estimates. Theoretically, however, the posterior distributions encompass all values, i.e., from minus infinity to plus infinity. For most of these values, the density function is near zero and they are extremely unlikely to occur. If we run an extreme amount of iterations, however, the thin tails of the posterior distributions will also be populated to some extent. Some of these unlikely values may result in convergence problems. For example, a variance covariance matrix in Mplus is considered singular when the condition number of the matrix is less than 10^{-10}. With a vary small probability that can occur even for a perfectly good model. If we run an excessive amount of iterations, it is possible to reach these unlikely points in the posterior distribution which will cause the computation to stop. We do not recognize this as a problem with the model or the estimation but with the setting requiring the excessive number of iterations. Instead of using an excessive number of iterations, it is more appropriate to use the Mplus default convergence criterion and the automated determination for the number of MCMC iterations. Before accepting the model as final, we also recommend doubling the number of iterations. If the PSR convergence criterion continues to be satisfied in all of the additional iteration, we conclude that the model has converged and the number of iterations is sufficient to populate the main part of the posterior distribution.

Generally, when the number of iterations is above 50000, it could be considered excessive. Such number of iterations should be used only if it is necessary to establish convergence, i.e., if the PSR criterion continuous to exceed 1.1. Also, an excessive number of iterations can occur via the use of the THIN option in Mplus. The THIN command simply hides the intermittent draws, but can equally well lead to the perils of the extreme tails of the posterior distribution.

2.12 Factors with one or two indicators

Factors with one or two indicators are generally difficult to identify and may cause convergence problems even without any auto-regressive structures. In principle, such models can be estimated in DSEM, however, such estimation often requires very rich data sets that are rare in real applications. The solution here is to limit the flexibility of the model. For example, a two-

indicator factor model would be well identified if both loadings are fixed to 1.

2.13 Lack of variation in the data

If variables do not change within a cluster it is difficult to see how one can construct a meaningful time-series model for such variables. Mplus will list all variables that do not change within clusters as warning messages. The clusters can be removed from the analysis using the VARIABLE command option USEOBSERVATIONS = Y EQ _CLUSVAR.

In principle, information from the entire population can be used by the estimation to make inference for a cluster with no variation/information, but such an approach is limited in scope. It depends on how many of these clusters are in the population. The more non-informative data occurs in the population the more likely it is that the estimation will fail. Holding auto-regressive parameters invariant across clusters, i.e., using non-random auto-regressive parameters instead of random, can be used to resolve such convergence problems. Alternatively, removing clusters with such low level of information from the analysis all together can improve the model estimation. Because the amount of information contained in such cluster is very small, removing these data points from the population is unlikely to affect the overall analysis dramatically.

2.14 Categorical variables that are predominately constant

Time-series model estimation is informed by how variables change across time. If the variables do not change much across time, it will be very difficult to build time-series models. If a binary variable remains mostly zero across time, it will be difficult to build a well-identified subject-specific time-series model, unless the time series data is very long. For categorical data that is predominately constant, we recommend using population level auto-regressive parameters, i.e., non-random auto-regressive parameters.

2.15 Using priors to resolve convergence problems

In some situations, convergence problems can be resolved by specifying weakly informative or proper priors instead of using the Mplus default uninforma-

tive and improper priors. Priors can help the estimation when the number of clusters is small. The most common scenario when priors can resolve convergence problems is when the variance covariance matrix on the between level is singular. The default prior for a variance covariance matrix in Mplus is $IW(0, -p-1)$ where p is the size of the variance covariance matrix. One alternative prior specification is $IW(I, p+1)$ where I is the identity matrix. The advantage of this prior is that it is a proper prior and it provides uninformative/uniform marginal prior for the correlation parameters, see Asparouhov and Muthén (2010) and Asparouhov and Muthén (2021). Using such proper prior for the between level variance covariance matrix can resolve convergence problems in some situations. Consider the bivariate DSEM example given in Figure 7. Using the Mplus default priors, withe the model prior statement commented out, all 5 replications fail to converge. Using the prior $IW(I, p+1)$, withe the model prior statement, only 1 of the replications fails to converge.

Figure 7: Using priors to resolve convergence problems

```
MONTECARLO:  NAMES ARE y z;
NOBS = 800; NREP = 5;
NCSIZES = 2; CSIZES = 30(20) 20(10);
lagged = y(1) z(1);
missing=all;

model missing: [y*0 z*0];

ANALYSIS:   TYPE = TWOLEVEL RANDOM;
estimator=bayes;  proc=2; biter=(30000);

MODEL POPULATION:

  %WITHIN%
  y on y&1*0.5;
  z on z&1*0.5;
  z on y&1*0.2;
  y on z&1*0.2;
  y*0.5 z*0.5; y with z*-0.25;

  %BETWEEN%
  y*1; z*1; y with z*0.6;

MODEL:

  %WITHIN%
  y on y&1*0.5;
  z on z&1*0.5;
  z on y&1*0.2;
  y on z&1*0.2;
  y*0.5 z*0.5; y with z*-0.25;

  %BETWEEN%
  y*1 (v1); z*1 (v2); y with z*0.6 (c);

!model prior:
!v1~IW(1,3);
!v2~IW(1,3);
!c~IW(0,3);
```

3 Building up the model

A two-level DSEM model should always be constructed in stages. Start with the simplest model to the most advanced. Using this strategy will reduce the likelihood of an unidentifiable convergence problem. We recommend using the following estimation sequence.

1. **Two-level model**
   This model can simply be the unrestricted/unstructured two-level model. In this model all variables are correlated on the within and the between levels, but there are no lagged variables involved in the model at all. Alternatively, the model can be the structured two-level model that has the structure of the DSEM model without the auto-regressive parameters.

2. **Two-level DSEM model with random intercepts only**
   All other parameters should be non-random. In Asparouhov et al. (2018) and Asparouhov and Muthén (2020) it is shown that the RDSEM model is the more natural extension(transition) from the two-level model than the DSEM model. Note that the RDSEM and the DSEM models differ only if there are covariates in the model. Even if the DSEM model is the ultimate goal, it might be useful at this stage of the analysis to explore the RDSEM model as well. This is easy to do in Mplus and involves replacing the & symbol (used for lagged regressions) with the ^ symbol.

3. **Two-level DSEM model with uncorrelated random intercepts and slopes**
   Random slopes should be added one at a time. Random slopes should be uncorrelated and unstructured on the between level. Random effects that have non-significant variance (Z score less than 3) should be converted back to non-random.

4. **Two-level DSEM model with correlated random intercepts and slopes**
   Random effects can be structured or correlated on the between level at this stage.

5. **Two-level DSEM model with random intercepts, slopes and variances**

6. **Two-level DSEM model with random intercepts, slopes, variances and covariances**

The simpler the model is, the easier it is to correctly identify the source of a convergence problem or estimation instability. Gradually building up the model will allow us to quickly identify the model component that caused the problem. That would be the last component that was added to the model.

We recommend that variables are combined gradually. First, a univariate auto-regressive model should be estimated for all the variables of interest. These univariate models can then be combined into bivariate and multivariate auto-regressive models. Hamaker et al. (2021) illustrates in detail such a build up sequence of models. In addition, Hamaker et al. (2021) recommends analyzing the time-series data one person at a time as a preliminary exploration step.

4 Conclusion

The DSEM modeling framework is fairly complex and a variety of phenomenons discussed in this article can complicate the analysis and result in estimation problems. Estimating a SEM model, a multilevel regression model, or a time-series model is much simpler than estimating a DSEM model which combines all of these modeling frameworks in one. Approaching the DSEM analysis as if it is a simple regression is going to be futile. The analysis requires plenty of preliminary analysis. If the final model that is desired is attempted as the very first model estimation, it is very likely that estimation problems will occur.

In addition to the estimation complexities of the DSEM model, the analysis is complicated by a variety of data related problems that are specific to intensive longitudinal data (ILD) encountered in practice. It is extremely unusual in multilevel modeling applications to have clusters with very few or even a single observation. ILD data on the other hand often includes such situations. It is extremely unusual in multilevel modeling applications to have all observations within a clusters be identical. For ILD data, this is a fairly common situation. It is quite unusual in multilevel modeling applications to have 80% missing data. For ILD data, this is a fairly common situation because often the times of observation are randomized and are different from person to person. Missing data results from the fact that observations are not recorded in every period.

These data related problems most likely can not be resolved entirely by the final analysis. It would be beneficial if the design of the data collection takes into account the model the data will be analyzed with. For example, the times of observations are often randomized to reduce expectation bias of the responders. It is often hypothesized that subjects respond differently if they know in advance when they need to record observations, i.e., it is hypothesized that the subject response can be inadvertently biased. If such data are to be analyzed with a VAR model, however, where we study the effects of observations in one period on the observations in the next period, collecting data with clearly identified concept of period would definitely enhance the quality of the study. Thus, randomized times of observations should be limited to being randomized within that concept of period to avoid creating large amounts of missing data.

References

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

[2] Asparouhov, T. & Muthén, B. (2010). Bayesian analysis using Mplus: Technical implementation. Technical appendix. Los Angeles: Muthén & Muthén http://www.statmodel.com/download/Bayes3.pdf

[3] Asparouhov, T., & Muthén, B. (2020). Comparison of models for the analysis of intensive longitudinal data. Structural Equation Modeling: A Multidisciplinary Journal, 27 , 275-297.

[4] Asparouhov, T. & Muthén, B. (2021). Bayesian analysis of latent variable models using Mplus. Technical report. Los Angeles: Muthén & Muthén. http://www.statmodel.com/download/BayesAdvantages18.pdf

[5] Hamaker, E. L., Asparouhov, T., Brose, A., Schmiedek, F., & Muthén, B. (2018). At the frontiers of modeling intensive longitudinal data: Dynamic structural equation models for the affective measurements from the COGITO study. Multivariate Behavioral Research, 53, 820–841.

[6] Hamaker, E.L., Asparouhov, T, & Muthén, B. (2021). Dynamic structural equation modeling as a combination of time series modeling, multilevel modeling, and structural equation modeling. To be published as Chapter 31 in: The Handbook of Structural Equation Modeling (2nd edition); Rick H. Hoyle (Ed.); Publisher: Guilford Press.