General Random Effect Latent Variable Modeling: Random Subjects, Items, Contexts, and Parameters

Tihomir Asparouhov and Bengt Muthén

November 19, 2014

Abstract

Bayesian methodology can be used to estimate cluster specific structural equation models with two-level data where all measurement and structural coefficients, including intercepts, factor loadings and regression coefficients can be estimated as cluster level random effects rather than fixed parameters. Bayesian methodology is also well-suited for estimating latent variable models where subjects are not the only random mode, but also items and contexts. A general cross-classified structural equation model is presented where observations are nested within two independent clustering variables. The models include continuous and categorical dependent variables. Various applications are discussed. The random loading model is used for estimating multiple group factor analysis models with large number of groups and measurement non-invariance. Individual differences factor analysis model is described where factor loadings, as well as factor means and variances are individually specific. This model is used for analyzing ecological momentary assessment data for mood disorders. Finally, the cross-classified structural framework is used for the estimation of time intensive longitudinal structural models.

1 Introduction

In this article we describe some new modeling possibilities for multilevel and cross classified data. Using Bayesian methodology it is possible to substantially expand the set of structural equation models beyond the set of

models that can be estimated with maximum likelihood estimation (MLE) or weighted least squares (WLS) estimation methodology. Currently the latter two estimators provide the basis for almost all structural equation model fitting.

The first modeling extension described in this article is for two-level SEM where the measurement loadings are allowed to vary across clusters. The standard two-level SEM, which can be estimated with the ML estimator, has cluster specific random intercepts as well as cluster specific regression slopes for observed predictors. If the predictor variable however is an unobserved latent factor as in the case of the factor analysis measurement model the slope coefficient has to be a fixed parameter otherwise the MLE will require numerical integration. In practical settings this is not feasible when the number of cluster specific effects is greater than 3 or 4. In the case of categorical dependent variables WLS estimation can be used to estimate two-level SEM with random intercepts, see Asparouhov and Muthén (2007), however the WLS method can not estimate models with random slopes even for observed variables. The MLE estimator can be used with categorical variables to estimate models with random intercepts, slopes, and loadings, however again the method is not feasible when the number of random effects is greater than 3 or 4. In this article we show that the Bayesian methodology can be used to easily estimate SEM with any number of random intercepts, slopes and loadings with categorical and continuous dependent variables. This is a breakthrough not only because random loadings are included in this more general framework but also because models with large number of cluster specific random effects can now be easily estimated without encountering inadmissible solutions. The ML and WLS estimators frequently yield non-positive definite or singular variance covariance matrices, or negative residual variances when the number of random effects is larger. The Bayesian methodology does not have this flaw.

Cross classified modeling has been of interest in the last two decades as the next step of hierarchical modeling, see for example Fielding and Goldstein (2006) for an overview. Cross classified models are used to model hierarchical data which has two-way clustering. For example students are clustered within schools but are also cross clustered within neighborhoods. Both clustering levels can have an effect on the individual outcomes. Until recently most of the research and applications of cross classified modeling has been focused on regression models. The Bayesian methodology has been used in Browne et al. (2007) and Gonzalez et al. (2008) to estimate multivariate and SEM cross-

classified models. In this article, we describe a generalization of the cross-classified SEM model described in Gonzalez et al. (2008). We extend the model to include structural equations on all three levels: individual level and the two clustering levels, random slopes and loadings as well as interactions between random effects from the two different clustering levels.

It is clear that a multitude of recently proposed Bayesian methods are available for psychological research but are largely unknown to psychologists. Promising new methods extensions are ready for exploration by psychological researchers. Furthermore, it has not been made sufficiently clear how to use the new methods in practice. As an attempt to improve the situation, this paper has three aims. The first is to give an overview of the many new and promising Bayesian approaches to random effect latent variable modeling that have been proposed in the recent statistical literature. The second is to make several additional modeling contributions. The third is to place all these new methods in the general latent variable modeling framework of the Mplus program as implemented in Mplus Version 7. Mplus input for all analysis are available at www.statmodel.com. In Section 2 we describe the random loadings model. In Section 3 we describe the general cross-classified SEM. In Section 4 we discuss model fit evaluation for the two-level and cross-classified models. In Section 5 we illustrate the random loading model with a factor analysis model on teacher effectiveness. In Section 6 we extend the random loading model to include cluster specific factor mean and variance parameters. In Section 7 we define the individual differences factor analysis model and use that model for analyzing ecological momentary assessment data for mood disorders. In Section 8 we use the general cross-classified structural equation modeling framework for the estimation of time intensive longitudinal structural models on students aggressive-disruptive behavior. Section 9 concludes.

2 Two-level SEM with random loadings

In this section we describe a generalized two-level SEM with random loadings, i.e., the factor loadings in this model are cluster specific random effects. Let Y_{ij} be the vector of observed variables. The general two-level structural equation model, as described in Muthén (1994), is defined by the following equations

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

$$Y_{1,ij} = \Lambda_1 \eta_{1,ij} + \varepsilon_{1,ij} \tag{2}$$

$$\eta_{1,ij} = B_1 \eta_{1,ij} + \Gamma_1 x_{1,ij} + \xi_{1,ij} \tag{3}$$

$$Y_{2,j} = \nu_2 + \Lambda_2 \eta_{2,j} + \varepsilon_{2,j} \tag{4}$$

$$\eta_{2,j} = \alpha_2 + B_2 \eta_{2,j} + \Gamma_2 x_{2,j} + \xi_{2,j}. \tag{5}$$

The variables $x_{1,ij}$ and $x_{2,j}$ are the vectors of predictor variables observed at the two levels. The variables $\eta_{1,ij}$ and $\eta_{2,j}$ are the vectors of latent variables. The residual variables $\varepsilon_{1,ij}, \xi_{1,ij}, \varepsilon_{2,j}, \xi_{2,j}$ are zero mean normally distributed residuals with variance covariance matrices $\Theta_1, \Psi_1, \Theta_2, \Psi_2$ respectively. The parameters $\alpha_k, \Lambda_k, B_k, \Gamma_k$, $k = 1, 2$, are the model parameters to be estimated. The above model can easily be extended to categorical variable using the probit link function. For each categorical variable Y_{pij} we define the underlying variable Y^*_{pij} such that

$$Y_{pij} = l \Leftrightarrow \tau_{l-1,p} < Y^*_{pij} < \tau_{lp} \tag{6}$$

where τ_{lp} are the threshold parameters. To include a categorical variable in the general model above we substitute the variable Y_{pij} with Y^*_{pij}. For identification purposes the variance of $\varepsilon_{1,pij}$ is fixed to 1. The two-level model described above is the standard two-level SEM model. We extend this model to allow the parameters Λ_1, B_1, and Γ_1 to be cluster specific random effects instead of fixed parameters. These random effects are essentially between level latent variables, i.e., are variables in the vector $\eta_{2,j}$.

The Bayesian estimation of this model is a marginal extension of the standard two-level SEM estimation as described in Asparouhov and Muthén (2010a) and (2010b). Conditional on the within level latent variable the Gibbs sampler for the between level random effects is the same as if the within level latent variable is observed, i.e., once the within level latent variable is sampled the conditional posteriors are as in the regular SEM model with random effects on observed variables. More details on the Bayesian estimation of the above model can be found in De Jong et al. (2007) and De Jong and Steenkamp (2009).

The random loadings technique has a number of practical applications. For example, in multilevel structural equation models allowing the random loadings to vary across clusters yields a more flexible model. Another practical application is in the case of multiple group SEM when there are many groups and many minor but significant loadings differences between the groups, see De Jong et al. (2007) and Davidov et al. (2012) for modeling cross-cultural measurement non-invariance. In that case the random loadings approach can be used to obtain a better fitting and a more parsimonious model that allows for loadings non-invariance between the groups without increasing the number of parameters. In addition, the random loadings model would avoid the step-wise modeling strategy that relaxes loading equalities one at a time.

3 Cross-classified SEM

Cross-classified data arises in various practical applications. For example, in a longitudinal study the students performance scores are clustered within students and within teachers, see Luo and Kwok (2012). Such a model can be used to simultaneously estimate the students innate abilities and the teachers abilities and can be combined with a student level growth model. Another example arises in survey sampling where observations are nested within neighborhoods and interviewers.

A different type of cross-classified applications arises in the analysis of multiple random mode data as discussed in Gonzalez et al. (2008), where observations are nested within persons and cross nested within treatments or situations. In multiple random mode data, there are two or more dimensions that vary randomly. For example, one random dimension can be persons, while the second random dimension can be situational contexts. Persons are random samples from a population. Contexts are also random samples from a target population of contexts. Observations are collected for various persons in various contexts. Experimental design data can also be treated as multiple random mode data when various experiments are conducted for various subjects. The experiments are randomly selected from a large population of possible experiments and subjects are randomly selected from a population as well.

Finally, a third type of cross-classified applications arises in Generalizability theory, see Cronbach et al. (1963), Marcoulides (1999), Brennan (2001)

and De Boeck (2008) where test questions are modeled as random samples from a population of test items. In this case observations are nested within persons and also cross-nested within items.

It should be noted the importance of accounting for all sources of clustering when analyzing clustered data. If one of the clustering effects is ignored, the model is essentially underspecified and fails to discover the true explanatory effect stemming from the additional clustering. This misspecification can also lead to underestimating or overestimating of the standard errors, see Luo and Kwok (2009).

While basic multilevel models are easy to estimate through maximum-likelihood via the EM algorithm, see Raudenbush and Bryk (2002) and Goldstein (2011), the cross classified models are not. Rasbash and Goldstein (1994) found a way to obtain the maximum-likelihood estimates for a cross classified model by respecifying the model as a multilevel model. However that method can not be used in general settings when the number of cluster units at both classification levels is large. The MCMC estimation method proposed in Browne et al. (2001) is a general estimation method for fitting cross classified models with no restriction on the data structure. The method can be applied for normally distributed variables as well as categorical variables.

Suppose that the observed data is clustered within J level 2 units. Suppose also that there is a different type of clustering for this data, we will call this level 3 clustering, with K level 3 units. The data is cross classified when the level 2 clusters are not nested within the level 3 clusters and vice versa. Denote by Y_{pijk} the $p-$th observed variable for person i belonging to level 2 cluster j and level 3 cluster k. The number of observations that belong to level 2 cluster j and level 3 cluster k can be any number including 0. In some special experimental designs the number of such observations is exactly 1, see for example Gonzalez et al. (2008). Raudenbush (1993) shows that the cross classified modeling can be conducted even for sparse situations in which most of the intersection cells (j,k) are empty.

Denote by Y_{ijk} the vector of all observed variables. The definition of the cross-classified SEM model is based on the following 2-way ANOVA equation

$$Y_{ijk} = Y_{1,ijk} + Y_{2,j} + Y_{3,k} \tag{7}$$

where $Y_{2,j}$ is the random effect contribution of the j-th level 2 cluster, $Y_{3,k}$ is the random effect contribution of the k-th level 3 cluster and $Y_{1,ijk}$ is the

individual level residual. Alternative interpretation for $Y_{2,j}$ and $Y_{3,k}$ is that they are the random intercepts in the linear model.

At this point we can define separate structural equations for the 3 sets of variables at the 3 different levels

$$Y_{1,ijk} = \nu_1 + \Lambda_1 \eta_{1,ijk} + \varepsilon_{1,ijk} \tag{8}$$

$$\eta_{1,ijk} = \alpha_1 + B_1 \eta_{1,ijk} + \Gamma_1 x_{1,ijk} + \xi_{1,ijk} \tag{9}$$

$$Y_{2,j} = \Lambda_2 \eta_{2,j} + \varepsilon_{2,j} \tag{10}$$

$$\eta_{2,j} = \alpha_2 + B_2 \eta_{2,j} + \Gamma_2 x_{2,j} + \xi_{2,j} \tag{11}$$

$$Y_{3,k} = \Lambda_3 \eta_{3,k} + \varepsilon_{3,k} \tag{12}$$

$$\eta_{3,k} = \alpha_3 + B_3 \eta_{3,k} + \Gamma_3 x_{3,k} + \xi_{3,k} \tag{13}$$

The variables $x_{1,ijk}$, $x_{2,j}$ and $x_{3,k}$ are the vectors of predictor at the three different levels. The variables $\eta_{1,ijk}$, $\eta_{2,j}$ and $\eta_{3,k}$ are the vectors of the latent variables on the three different levels. The residual variables $\varepsilon_{1,ijk}, \xi_{1,ijk}, \varepsilon_{2,j}, \xi_{2,j}, \varepsilon_{3,k}, \xi_{3,k}$ are zero mean normally distributed residuals with variance covariance matrices $\Theta_1, \Psi_1, \Theta_2, \Psi_2, \Theta_3, \Psi_3$ respectively. In the above equations the intercept parameter vector ν_1 can be present just in one of the three levels. The parameters $\alpha_k, \Lambda_k, B_k, \Gamma_k$, $k = 1, 2, 3$, are fixed model parameters, however, Λ_1, B_1, Γ_1 can be cluster specific random effects defined at cluster level 2 or cluster level 3 or it can be the sum of two random effects, one defined at level 2 and one defined at level 3. In addition, Λ_2, B_2, Γ_2 can be cluster specific random effects defined at cluster level 3 which will allow the interaction modeling between observed and unobserved variables defined on the two clustering levels. These random effects are essentially between level latent variables, i.e., are variables in the vectors $\eta_{2,j}$ or $\eta_{3,k}$. Categorical data can easily be incorporated in this model through the probit link approach used in the previous section.

The estimation of the above model is based on the MCMC algorithm with the Gibbs sampler and is only marginally different from the estimation of the two-level structural equation model described in Asparouhov and Muthén (2010a) and Asparouhov and Muthén (2010b). The two-level model estimation is based on first sampling the between component for each variable from its posterior distribution. Then the Gibbs sampler for two-group

structural equation models is used to sample the rest of the components. The two groups are the within and the between levels. Similarly here we sample the two between level components in separate Gibbs sampling steps

$$[Y_{2,j}|*, Y_{ijk}, Y_{3,k}] \tag{14}$$

and

$$[Y_{3,k}|*, Y_{ijk}, Y_{2,j}]. \tag{15}$$

Both of these posterior distributions are obtained the same way as the posterior for the two-level between components since conditional on $Y_{3,k}$ the model for Y_{ijk} is essentially a two-level model with the between component being $Y_{2,j}$. Similarly the conditional posterior distribution of $Y_{3,k}$ given $Y_{2,j}$ is the same as the posterior for the two-level component in a two-level model. After the above two steps, the Gibbs sampler continues sampling the remaining components as if the model is a 3 group structural equation model where $Y_{1,ijk}$, $Y_{2,j}$ and $Y_{3,k}$ are the observed variables in the 3 groups.

4 Model fit

Model fit can be evaluated in two different ways, using the PPP (posterior predictive p-value), see Meng (1994), and using DIC, see Spiegelhalter et al. (2002). These inference statistics are available under certain conditions. The PPP is available when there are no random slopes and loadings. DIC is available when all the variables are continuous.

First we describe the PPP fit statistic. The PPP value is based on a test statistic f. For a cross-classified model the test statistic we use is as follows

$$f(Y_{1,ijk}, Y_{2,j}, Y_{3,k}, \theta) = f_1(Y_{1,ijk}, \theta) + f_2(Y_{2,j}, \theta) + f_3(Y_{3,k}, \theta) \tag{16}$$

where θ represents the model parameters. The function $f_1(Y_{1,ijk}, \theta)$ is the standard chi-square SEM test of fit which tests the model implied level 1 mean and variance covariance against the unrestricted mean and variance covariance on level 1. Similarly $f_2(Y_{2,j}, \theta)$ and $f_3(Y_{3,k}, \theta)$ are the chi-square tests of fit for level 2 and 3. The posterior predictive p-value is the proportions of MCMC iterations for which

$$f(Y_{1,ijk}, Y_{2,j}, Y_{3,k}, \theta) < f(Y^*_{1,ijk}, Y^*_{2,j}, Y^*_{3,k}, \theta) \tag{17}$$

where $Y^*_{1,ijk}$, $Y^*_{2,j}$ and $Y^*_{3,k}$ are model generated data using the current MCMC generated parameters θ. This PPP value can be used similarly to the frequentist p-values. A value above 0.05 or 0.01 indicates a good model fit while a value smaller than these cut off levels indicates poor model fit. The PPP can be used as a test of fit for a structural cross-classified model. The structural model is essentially tested against the unrestricted cross-classified model where the variance covariance on each level is unrestricted. For two-level structural models the PPP value is computed similarly but the test function f consists of only two chi-square statistics coming from the two different levels.

To compute DIC for the cross-classified and the two-level models we compute the deviance $D(\theta)$ at each MCMC iteration

$$D(\theta) = -2\log(p(Y|\theta)) \tag{18}$$

where θ represents all model parameters as well as all between level random effects. We then compute the effective number of parameters p_D and DIC

$$p_D = \bar{D} - D(\bar{\theta}) \tag{19}$$

$$DIC = p_D + \bar{D} \tag{20}$$

Note that p_D is the only penalty for model complexity with this information criterion. DIC can be used to compare any number of competing models, not necessarily nested. The best model is the model with the lowest DIC value. Note that the definition of DIC is somewhat ambivalent. In single level models, for the computation of the deviance we condition only on the model parameters and not on any latent variables or random effects. Latent variables and random effects are marginalized. In two-level and cross-classified models we condition also on all the between level random effects. Some models can be estimated as single level wide models or as two-level long models which leads to two different definitions of DIC for the same model. In using DIC for comparison it is important that the DIC is computed in the same framework, i.e., it would be incorrect to compare wide model DIC with long model DIC. In single level models DIC is asymptotically the same as AIC computed with maximum-likelihood estimation. This property does not hold for two-level and cross-classified models again due to the fact that random effects are not marginalized. In single level models the effective number of parameters p_D is approximately the same as the number of model parameters.

In two-level and cross-classified models the effective number of parameters also includes all between level random effects, i.e., for each random effect in the model p_D is expected to increase by the number of clusters where the random effect is defined. If there are 100 clusters in the data, each between level random effect will contribute approximately 100 to p_D as that is the number of random parameters that are used to fit the model. However, the more correlated the random effects are the lower the p_D contribution is, i.e., if two random effects are very highly correlated they will not contribute as two parameters but possibly as one and a half. Also the smaller the random effect variance is the smaller the p_D contribution is. These observations are important as p_D is the only complexity penalty in the DIC criterion. The estimation of DIC appears to need more MCMC iterations than the estimation of the model parameters. This is particularly the case when the model has a large number of random effects. Therefore if DIC is used for model comparison a large number of MCMC iterations should be used for the estimation. Similar to AIC, DIC does not penalize sufficiently over-parameterized models. While BIC asymptotically selects the true model, that is not true for AIC and DIC. AIC and DIC may not select the true model but will always select a well fitting model, possibly somewhat over-parameterized. AIC and DIC are also more robust than BIC. If the true model is not among those that are compared AIC and DIC are expected to perform better than BIC. In the presence of missing data there are also two ways of computing DIC. The missing data can be marginalized or it can be treated as yet additional parameters. These two-versions of computing DIC are also not comparable. In model selection one of the two versions should be used for all competing models. In Mplus the missing data is marginalized and does not contribute to the estimated number of parameters while in MLwiN for example the missing data are treated as additional parameters.

Next we illustrate how DIC can be used for model comparison. We generate a data set according to a cross-classified factor analysis model with five continuous indicators and one factor on each of the levels. All loadings and residual variance parameters are set to 1 and all intercept parameters are set to 0. We analyze the data using the true cross-classified model, which we denote by model M1, as well as a two-level factor analysis model M2 where one of the two clustering variables is ignored. The cross-classified data set has 100 clusters on each level with just one observation in each pair of crossed clusters for a total of 10000 observations. The DIC results are presented in Table 1 and clearly show that model M1 is better than model M2 by a wide

Table 1: DIC simulation results for cross-classified factor analysis data

Model	p_D	DIC
M1	996.6	161084
M2	499.7	196113

Table 2: DIC simulation results for two-level factor analysis data

Model	p_D	DIC
M1	558.1	160380
M2	504.2	160373

margin. The estimated number of parameters p_D matches the number of random effects for the two models: 1000 random effects for model M1 and 500 random effects for M2.

Next we generate a data set according to model M2 and analyze the data again using model M1 and M2, i.e., the data is a two-level data and we analyze it with a two-level model as well as a cross-classified model where the extra clustering level doesn't have any effect on the data. As expected the random effects for model M1 on the extra level are estimated to be very close to zero and their variances are close to zero. These DIC results are presented in Table 2. Only by a small margin DIC prefers the true model M2. The deviance D for the two models would be very close as both estimated models imply the same, namely, that the data should be analyzed as two-level data and the effect of the extra clustering is near zero. Model M2 obtained better DIC due to having smaller effective number of parameters. For model M2 the effective number of parameters matches the 500 random effects. For model M1 the effective number of parameters is 558 which is far off the 1000 random effects in the model. The reason that happens is because the random effects on the extra level are barely detectable as they are near zero. Nevertheless model M2 is assessed to have larger number of effective parameters and thus DIC prefers the correct model M1.

5 Student evaluation of teacher effectiveness example

In this section we describe a practical application of the random loading model. We use the SEEQ data, student evaluation of teacher effectiveness, described in Marsh and Hocevar (1991). The data consists of 35 continuous items. The teacher evaluations are split in 21 subsamples based on the qualifications of the teacher and the academic discipline. In Marsh and Hocevar (1991) a 9 factor analysis model is considered. For simplicity however, we consider a 1-factor analysis model, although using several factors does not elevate the complexity of the model particularly when all measurements load on a single factor. It was noted in Marsh and Hocevar (1991) that minor variation exists in the loadings between the 21 groups. The sample size in this application is 24158 and therefore any minor variations in the factor loadings between the groups would be statistically significant. Thus the model that imposes measurement invariance would be rejected. Such sample size and group combinations are not unusual. Davidov et al. (2012) analyzes data from 26 countries and 43779 observations.

The model without measurement invariance would have more than 1500 parameters and thus would not be parsimonious. It is possible to evaluate the measurement non-invariance for each variable and group however that would be a very tedious process given that over 1500 parameters are involved and it is not clear which subset of parameters should be held equal and which should not be held equal. If the measurement non-invariance is ignored the factor score estimates which represent the teacher effectiveness could have a substantial error. One method for dealing with measurement non-invariance using the Bayesian methodology is developed in Muthén and Asparouhov (2012) where all parameters are estimated but approximate equality is enforced between the loadings across groups, i.e., the loadings are estimated as different parameters but a strong prior specification is used that amounts to holding the loadings approximately equal. This approach treats the loadings as fixed parameters rather than random effects. In this section we will illustrate how the random loading model can be used to easily resolve the measurement invariance problems.

To illustrate the random loading model we standardize all variables and estimate three different models. Let Y_{ij} be the observed indicator vector for individual i in group j. The first model M1 is the standard 1-factor analysis

model estimated with the maximum-likelihood estimator

$$Y_{ij} = \alpha + \lambda\eta_{ij} + \varepsilon_{ij} \tag{21}$$

where $\eta_{ij} \sim N(0,1)$ and $\varepsilon_{ij} \sim N(0,\Theta)$.

The second model M2 is a 1-factor analysis model with random intercepts estimated with the maximum-likelihood estimator

$$Y_{ij} = \alpha_j + \lambda\eta_{ij} + \varepsilon_{ij} \tag{22}$$

where $\eta_{ij} \sim N(0,1)$, $\varepsilon_{ij} \sim N(0,\Theta)$, and $\alpha_j \sim N(\alpha,\Sigma)$.

The third model M3 is a 1-factor analysis model with random intercepts and loadings using the Bayes estimator

$$Y_{ij} = \alpha_j + \lambda_j\eta_{ij} + \varepsilon_{ij} \tag{23}$$

where $\eta_{ij} \sim N(0,1)$, $\varepsilon_{ij} \sim N(0,\Theta)$, $\alpha_j \sim N(\alpha,\Sigma)$, and $\lambda_j \sim N(\lambda,\Sigma_2)$. In Models M2 and M3 we estimate diagonal variance covariance matrices Σ and Σ_2 for simplicity.

Table 3 contains the factor loading estimates for the 35 variables and the three models. In Models M3 we also include the 95% range for the cluster specific factor loadings based on the estimated normal distribution for the random loadings. It can be seen from these results that the random loading range is quite wide and that cluster specific loadings can be substantially different from their fixed ML based estimates. The advantage of model M3 is that it does not assume measurement invariance of the factor analysis model and thus it is more flexible than models M1 and M2. Model M3 simply assumes that a one factor analysis model holds in each group without assuming loading or mean group invariance.

Table 4 compares the overall model fit using DIC for the three models. Model M3 clearly outperforms models M1 and M2 for this data. Also model M2 provides better fit than M1. The decrease in the DIC value obtained by allowing cluster specific means is about 7 times larger than the decrease in the DIC value obtained by allowing also cluster specific variance covariance. This quantification of the decrease might be useful in comparing different data sets. The effective number of parameters p_D for model M1 matches approximately the actual number of parameters 105. For model M2, p_D matches approximately the number of between level random effects 735 (21 clusters times 35 random intercepts) plus the within level parameters 70 for

Table 3: Factor loading estimates for SEEQ data

M1	M2	M3
0.76	0.79	0.81(0.67,0.94)
0.78	0.81	0.82(0.71,0.92)
0.80	0.80	0.82(0.71,0.93)
0.73	0.74	0.76(0.67,0.85)
0.88	0.83	0.85(0.70,1.01)
0.89	0.83	0.85(0.74,0.96)
0.84	0.75	0.77(0.64,0.91)
0.90	0.87	0.89(0.83,0.95)
0.81	0.85	0.87(0.73,1.01)
0.78	0.85	0.87(0.71,1.02)
0.78	0.83	0.84(0.71,0.98)
0.75	0.71	0.73(0.58,0.88)
0.80	0.66	0.67(0.48,0.85)
0.79	0.65	0.66(0.47,0.86)
0.86	0.80	0.82(0.68,0.95)
0.84	0.73	0.75(0.59,0.90)
0.73	0.75	0.77(0.57,0.97)
0.70	0.75	0.77(0.60,0.93)
0.76	0.77	0.80(0.66,0.93)
0.66	0.69	0.72(0.58,0.85)
0.82	0.76	0.78(0.64,0.92)
0.84	0.77	0.79(0.64,0.94)
0.86	0.78	0.80(0.65,0.95)
0.80	0.73	0.75(0.63,0.87)
0.70	0.77	0.79(0.66,0.91)
0.74	0.78	0.80(0.62,0.98)
0.73	0.78	0.79(0.65,0.93)
0.62	0.63	0.64(0.44,0.84)
0.68	0.73	0.74(0.57,0.90)
0.83	0.89	0.90(0.77,1.04)
0.92	0.92	0.93(0.84,1.02)
0.17	0.18	0.17(0.01,0.34)
0.18	0.16	0.15(-0.03,0.32)
0.11	0.14	0.15(-0.06,0.35)
0.14	0.11	0.10(-0.08,0.27)

Table 4: DIC comparison for factor loading models for SEEQ data

Model	p_D	DIC
M1	104	1594590
M2	792	1523405
M3	1443	1514784

a total of 805. For model M3, p_D matches approximately the number of between level random effects 1470 plus the within level parameters 35 for a total of 1505.

6 Estimating group specific factor mean and variance

For many practical applications the multiple group factor analysis model with random intercepts and loadings (23) is insufficient. In the standard multiple group factor analysis models the goal is to estimate group specific factor mean and variance parameters and to explain as much of the difference in the observed variables distribution across groups as differences in the factor distribution across groups. Since model (23) is so flexible that it already has group specific intercepts and loadings it is not possible to identify additionally the group specific factor mean and variance. The group specific factor mean is confounded with the group specific intercepts and it is not possible to separate the two. Factor mean variability across groups can be absorbed by the intercept variability across groups. Similarly, the group specific factor variance is confounded with group specific factor loadings and it is not possible to separate the two. Factor variance variability across groups can be absorbed by the loading variability across groups. In standard multiple group modeling the factor mean and variance parameters are identified by estimating an invariant measurement model, i.e., by holding equal the loadings and intercepts across groups. Thus in the modeling framework of model (23) where all loadings and intercepts are group specific there is no information in the data about the group specific factor mean and variance parameters and it is a matter of interpretation of what factor mean and variance parameters are, i.e., it is a matter of subjective choice similar to how in EFA the optimal

rotation is determined subjectively without any support from the data.

One simple way to identify the group specific factor mean and variance parameter is to assume that a particular factor indicator is invariant. Naturally the group specific factor mean and variance parameters will depend on which indicator is assumed to be invariant however the model fit will not. Regardless of which indicator is assumed to be invariant the model fit is as good as the fit of model (23) which essentially estimates a saturated group specific one factor analysis model. Other methods for estimating group specific factor mean and variance parameters are possible. In Fox (2010) for example the factor mean and variance are identified by constraining the intercept parameters to add up to zero in each group and the loading parameters to multiply to 1 in each group. In this section we provide a different approach to identifying the factor mean and variance that is consistent with two-level SEM methodology.

Consider first adding a group specific mean to model (22) as follows

$$\eta_{ij} = \eta_{2,j} + \eta_{1,ij} \tag{24}$$

where $\eta_{2,j} \sim N(0, \psi)$ and $\eta_{1,ij} \sim N(0, 1)$. Together with (22) this model can be reformulated as follows

$$Y_{ij} = Y_{2,j} + Y_{1,ij} \tag{25}$$

where

$$Y_{1,ij} = \lambda\eta_{1,ij} + \varepsilon_{ij} \tag{26}$$

$$Y_{2,j} = \lambda\eta_{2,j} + \alpha_j \tag{27}$$

The above equation shows that the group specific factor mean $\eta_{2,j}$ is identified as the factor behind the group specific indicator means $Y_{2,j}$, i.e., as the between level factor. Thus the factor mean is identified from the correlations between the random intercepts and is interpreted as the common change in the random intercepts.

Similarly we will show below that the factor variance can naturally be identified from the correlations between the random loadings and can be extracted from the common change of the random loadings. Consider the model

$$Y_{ij} = \alpha_j + \lambda_j\eta_{1,ij} + \lambda_j\eta_{2,j} + \varepsilon_{ij} \tag{28}$$

where α_j are the group specific mean $\alpha_j \sim N(\alpha, \Sigma)$ where Σ is a diagonal matrix, λ_j are the group specific loadings $\lambda_j \sim N(\lambda, \Sigma_2)$, $\eta_{2,j}$ is the group

specific factor mean $\eta_{2,j} \sim N(0, \psi)$, $\varepsilon_{ij} \sim N(0, \Theta)$ is the residual and the factor $\eta_{1,ij} \sim N(0, (1+\sigma_j)^2)$, has a group specific variance $(1+\sigma_j)^2$ where σ_j is a group specific random parameter $\sigma_j \sim N(0, \sigma)$. This model can be reformulated as follows

$$Y_{ij} = \alpha_j + \lambda_j(1+\sigma_j)\eta_{1,ij} + \lambda_j\eta_{2,j} + \varepsilon_{ij} \tag{29}$$

assuming $\eta_{1,ij} \sim N(0,1)$, i.e., the random loading is now $s_j = \lambda_j(1+\sigma_j)$. Denote by $\lambda_{j0} = \lambda_j - \lambda$ the residual of λ_j, i.e, $\lambda_{j0} \sim N(0, \Sigma_2)$. Now the random loading s_j is

$$s_j = \lambda_j(1+\sigma_j) = (\lambda + \lambda_{j0})(1+\sigma_j). \tag{30}$$

Under the assumption of approximate invariance, both λ_{j0} and σ_j are small and therefore their product $\lambda_{j0}\sigma_j$ is of smaller magnitude and it can be ignored. Therefore we obtain the following approximation

$$s_j \approx \lambda + \lambda_{j0} + \lambda\sigma_j. \tag{31}$$

The above equation shows that σ_j can be interpreted as the between level factor behind the random loadings, i.e., just as the the group specific factor mean can naturally be identified as the factor behind the random intercepts, the group specific factor variance can naturally be identified from the factor behind the random loadings.

From a practical perspective in multiple group modeling we want to have as much variation in the factor mean and variance and as little as possible in the intercept and the factor loadings to pursue measurement invariance or approximate measurement invariance. Factor analysis estimation tends to absorb most of the correlation between the indicators via the factors and to minimize the residual variances. Thus the model where the factor mean is the factor behind the random intercept and the factor variance is constructed from the factor behind the random loadings will try to explain as much as possible the variation between the correlation matrices across groups as a variation in the factor mean and variance rather than as a variation in the intercepts and the factor loadings. Thus this model is ideal for evaluating and separating factor means non-invariance from indicator intercepts non-invariance and factor variances non-invariance from factor loadings non-invariance. The model naturally separates the across-group variation in the factor mean and the across-group variation in the indicator intercepts. The

model also naturally separates the across-group variation in the factor loadings and the across-group variation in the factor variance. We conclude that the model is perfect for identifying factor mean and variance parameters in the presence of random intercepts and loadings. Other approaches to this identification problem use artificial constraints or unverifiable assumptions and therefore are inferior.

Note also that testing $\Sigma = 0$ is essentially a test for intercept invariance. Testing $\Sigma_2 = 0$ is a test for loading invariance. Testing $\sigma = 0$ is a test for factor variance invariance. Testing $\psi = 0$ is a test for factor mean invariance. In the Bayesian framework testing for significance of the variance of random effects has been described in Verhagen and Fox (2012) based on Bayes factor methodology. This method is also implemented in Mplus Version 7.

7 Individual differences factor analysis

To illustrate the model described in the previous section we use an example presented in Jahng et al. (2008)[1]. The data is obtained from an ongoing study of affective instability in borderline personality disorder (BPD) patients. Affective instability is considered a core feature of BPD that distinguishes this disorder from other disorders like depressive disorders. The data contains 84 individuals. Two groups of outpatients were entered into the study, 46 individuals with borderline personality disorder and 38 with major depressive disorder or dysthymic disorder. The mood factor for each individual was measured with 21 self-rated items on a scale of 1-5. We analyze the data assuming continuous distributions for the items and standardize the variables to have zero mean and variance one. For each individual the measurements were collected randomly several times a day over a 4 week period. In total between 76 to 186 assessments were conducted per person.

An analysis of the 21-item measurement instrument needs to take into account that repeated observations over time are correlated within individuals. One possible approach is the standard two-level factor analysis model where repeated measures are nested within individuals. Let Y_{pij} be the p-th item for individual i at assessment j and X_i be the binary indicator for the individual being classified with borderline personality disorder. Let P be the number of assessment items. A two-level factor model can be defined by the

[1]We thank Tim Trull and Phil Wood for providing the data and providing helpful comments

following two equations

$$Y_{pij} = \mu_p + \lambda_p \eta_{ij} + \zeta_{pi} + \varepsilon_{pij} \tag{32}$$

$$\eta_{ij} = \eta_i + \beta X_i + \xi_{ij}. \tag{33}$$

Here μ_p is the intercept parameter for the p-th item, ζ_{pi} is the individual variation from this mean parameter. The mean of ζ_{pi} is fixed to 0 for identification purposes and the variance parameter v_p is estimated. The factor η_{ij} is decomposed as the sum of the mean factor value η_i for individual i and the assessment specific deviation from that mean ξ_{ij}. The loading parameters λ_p are all estimated as well as the variance ψ of η_i while the variance of ξ_{ij} is fixed to 1 for identification purposes. The residual ε_{pij} is a zero mean normally distributed residual with variance θ_p. This model has $4P$ parameters: μ_p, λ_p, θ_p, and v_p as well as two parameters ψ and β for a total of $4P + 2$ parameters. Note again that with this approach the individual variable takes the role of the grouping variable because there are multiple assessments for each individual.

The two-level factor analysis model, however, has important shortcomings for these types of data. The model accommodates individually specific random intercepts for the factor as well as for each indicator through the latent variables η_i and ζ_{pi}. However, the model does not accommodate individual-specific factor variances, nor individual-specific factor loadings. The individual-specific factor variance is a key indicator of the individual's stability over time, in this case mood stability. Large factor variance for the mood factor is considered a core feature of BPD that distinguishes this disorder from other disorders like depressive disorders. Thus the individually specific factor variance is the most important feature in this study.

In addition, individual-specific factor loadings can capture differences in how individuals interpret the measurement instrument. For example, if one individual answers items 1 and 2 the same way and a second individual does not, then separate factor analysis models with individually specific factor loadings are needed to fit the data for both individuals. If the correlation matrix for the observed variables varies across individuals that means that the loadings should be individually specific. The data in this application is perfect to check whether the measurement instruments is interpreted the same way by different individuals because we have repeated measurements and we can easily estimate the individually specific parameters. This can

not be done in cross-sectional studies where each individual has just one measurement.

The individually specific factor variance is confounded with individually specific factor loadings. In the previous section we described a model that can naturally separate the across-individual variation in the factor loadings and the across-individual variation in the factor variance. We utilize this model in the BPD analysis. The model can be described as follows

$$Y_{pij} = \mu_p + s_{pi}\eta_{ij} + \zeta_{pi} + \varepsilon_{pij} \tag{34}$$

$$\eta_{ij} = \eta_i + \beta_1 X_i + \xi_{ij} \tag{35}$$

$$s_{pi} = \lambda_p + \lambda_p \sigma_i + \epsilon_{pi} \tag{36}$$

$$\sigma_i = \beta_2 X_i + \zeta_i. \tag{37}$$

Here we regress the factor and the variance of the factor on the BPD covariate X_i. The regression coefficient β_1 shows the amount of increase or decrease in the average factor value due to the borderline personality disorder. The regression coefficient β_2 shows the amount of increase or decrease of the factor variance due to the borderline personality disorder. Thus the two regression coefficients β_1 and β_2 represent the effect of the covariate X_i on the mean and the variance of the mood factor. The above model has an additional $P + 2$ parameters compared to model (32-33). These are the parameters β_2, $Var(\zeta_i) = \sigma^2$ and the P parameters $w_p = Var(\epsilon_{pi})$. An analysis based on this model will be called an Individual Differences Factor Analysis (IDFA). To summarize, the model accommodates individually specific factor intercept and variance, individually specific factor loadings as well as individually specific intercepts for each indicator variable.

It should be noted that when using IDFA individuals can be compared on their factor scores even when there is measurement non-invariance. The scores for the individual factor mean η_i in (35) can be estimated using Bayes plausible values.

The parameter estimates and standard errors for the IDFA model are presented in Table 5. There are 4 parameters not presented in Table 5: β_1, β_2, ψ and σ^2 and the estimates for these parameters are as follows: 0.702(0.116), 0.287(0.153), 0.584(0.107), 0.379(0.111) respectively. It is interesting to note that both β_1 and β_2 are statistically significant, i.e., the individuals with borderline personality disorder have significantly higher average mood factor

and higher variation in the mood factor compared to individuals with other major depressive disorder.

In Table 5 we also present the percentage of the factor loading variation that can be explained by the variation in the factor variance. That percentage varies from 25% to 81%. Testing the statistical significance for w_p using Verhagen and Fox (2012) method shows that all variance components are statistically significant, i.e., the loading parameters should be individually specific. This is a clear evidence that measurement instruments may not be interpreted the same way by different individuals and thus individual specific adjustments are needed to properly measure underlying factors. This fact is probably true even for many cross-sectional studies as well, however, individual specific adjustments to the factor loadings can be done only when repeated assessments are conducted.

Table 5: Individual Differences Factor Analysis: Borderline personality disorder study.

item	θ_p	λ_p	μ_p	v_p	w_p	Percentage loading invariance
Item 1	0.418(0.005)	0.214(0.026)	-0.221(0.058)	0.271(0.045)	0.042(0.007)	0.29
Item 2	0.400(0.005)	0.258(0.029)	-0.285(0.062)	0.317(0.054)	0.048(0.009)	0.34
Item 3	0.321(0.004)	0.465(0.044)	-0.449(0.030)	0.056(0.011)	0.025(0.006)	0.77
Item 4	0.314(0.004)	0.449(0.043)	-0.441(0.035)	0.089(0.016)	0.027(0.006)	0.74
Item 5	0.353(0.005)	0.442(0.045)	-0.429(0.019)	0.019(0.004)	0.061(0.011)	0.55
Item 6	0.390(0.005)	0.363(0.037)	-0.352(0.045)	0.152(0.026)	0.046(0.009)	0.52
Item 7	0.264(0.004)	0.390(0.037)	-0.373(0.051)	0.202(0.034)	0.020(0.004)	0.74
Item 8	0.253(0.003)	0.405(0.038)	-0.387(0.047)	0.176(0.029)	0.015(0.004)	0.81
Item 9	0.349(0.005)	0.345(0.037)	-0.341(0.033)	0.084(0.015)	0.054(0.010)	0.46
Item 10	0.240(0.003)	0.358(0.039)	-0.383(0.037)	0.102(0.019)	0.073(0.013)	0.40
Item 11	0.245(0.003)	0.383(0.036)	-0.361(0.053)	0.219(0.037)	0.019(0.004)	0.75
Item 12	0.337(0.005)	0.366(0.039)	-0.356(0.031)	0.074(0.014)	0.056(0.010)	0.48
Item 13	0.336(0.004)	0.319(0.032)	-0.309(0.054)	0.231(0.039)	0.030(0.006)	0.56
Item 14	0.211(0.003)	0.323(0.039)	-0.353(0.037)	0.102(0.018)	0.099(0.017)	0.29
Item 15	0.317(0.004)	0.314(0.033)	-0.323(0.065)	0.336(0.057)	0.047(0.009)	0.44
Item 16	0.307(0.004)	0.250(0.028)	-0.242(0.065)	0.339(0.057)	0.035(0.006)	0.40
Item 17	0.213(0.003)	0.348(0.039)	-0.369(0.042)	0.138(0.024)	0.074(0.013)	0.38
Item 18	0.343(0.005)	0.376(0.047)	-0.363(0.010)	0.002(0.001)	0.159(0.028)	0.25
Item 19	0.225(0.003)	0.391(0.049)	-0.390(0.016)	0.014(0.004)	0.174(0.029)	0.25
Item 20	0.279(0.004)	0.410(0.045)	-0.420(0.024)	0.038(0.008)	0.096(0.017)	0.40
Item 21	0.273(0.004)	0.408(0.048)	-0.390(0.015)	0.012(0.003)	0.153(0.026)	0.29

8 Intensive longitudinal data

Intensive longitudinal data have become quite common in recent years due to accumulating long-term longitudinal studies as well as a result of the use of new technological devices for data collection such as mobile devices, beepers and web interfaces, see Walls and Schafer (2006). New models need to be developed to analyze such data. Longitudinal data typically has been analyzed with multivariate models however if the number of time points is large these models can fail due to too many variables and parameters involved in the modeling. Estimating structural latent variable models in intense longitudinal settings can lead to additional challenges. Factor analysis models may be unstable over time and measurement invariance may be violated to some degree. Thus the time invariant structural models would be insufficient and inaccurate. The framework described in this article can resolve these problems. The random loading and intercept models can be used to model measurement and intercept non-invariance. These models have the advantage of borrowing information over time in the estimation the same way standard structural models do when assuming invariance. At the same time these new models have the advantage of accommodating measurement non-invariance the same way longitudinal structural models do. The models are also more parsimonious than longitudinal structural models. All these advantages will typically lead to more accurate model estimation.

To illustrate the intensive longitudinal modeling we will use the TOCA example described in Ialongo et al. (1999). The data consists of a teacher-rated measurement instrument capturing aggressive-disruptive behavior among a sample of U.S. students in Baltimore public schools. The instrument consists of 9 items scored as 0 (almost never) through 6 (almost always). A total of 1174 students are observed in 41 classrooms from Fall of Grade 1 through Grade 6 for a total of 8 time points. The multilevel (classroom) nature of the data is ignored in the current analysis. The item distribution is very skewed with a high percentage in the Almost Never category. The items are therefore dichotomized into the Almost Never versus all the other categories combined. For each student a 1-factor analysis model is estimated with the 9 items at each time point.

In the following sections we illustrate the three different approaches to intensive longitudinal data modeling: Longitudinal SEM, Multilevel SEM, Cross-classified SEM. We discuss the advantages and disadvantages of each method.

8.1 Longitudinal SEM

Let Y_{pit} be the $p-$th item for individual i at time t. The factor analysis model at time t is described by

$$P(Y_{pit} = 1) = \Phi(\lambda_{pt}\eta_{it} - \tau_{pt}) \tag{38}$$

where Φ is the standard normal distribution function resulting in a 2-parameter probit (normal ogive) IRT model. The variance of the aggressive behavior factor η_{it} is fixed to 1 and the mean is fixed to 0 for identification purposes. There are 8 times points and 9 items so in total we have 72 thresholds parameter τ_{pt} and 72 loading parameters λ_{pt}. In addition, the aggressive behavior factors η_{it} are nested within individual and thus we should account for the correlation between the individual factors across time. One approach is to simply estimate an unrestricted correlation matrix for the 8 individual factors. This would contribute an additional 28 correlation parameters for a total of 172 parameters. Because this model has 8 latent variables it cannot be estimated easily with the ML estimator which would require an 8 dimensional numerical integration. The model can be estimated however with the WLSMV estimator as implemented in the Mplus program, see Muthén and Muthén (1998-2014). The main problem with this model however is that it is not scalable in terms of time. The number of parameters grows as a quadratic function of the number of time points. In addition, the model is estimated as a multivariate model and this estimation is based on fitting an unrestricted multivariate probit model with 72 variables. This model has 72 threshold parameters and 2556 correlation parameters. The estimation of the unrestricted model can easily become computationally prohibitive as the number of time points increases. In addition, the sample size needed to estimate this unrestricted model may be substantial.

While the above model is flexible and accounts for measurement and threshold non-invariance it is difficult to provide interpretation for the variation in the parameters across time and to guarantee that the factor measurement model is sufficiently stable so that we can interpret it as the same factor changing over time. This leads to estimating additional models with ad-hoc parameter restrictions designed to parse those model parameters that are significantly different from those that are not. This process however is not feasible when the number of time points is substantial. A reasonable model to explore is the model that assumes complete measurement invariance

$$P(Y_{pit} = 1) = \Phi(\lambda_{p}\eta_{it} - \tau_{p}). \tag{39}$$

In addition we can impose a linear or quadratic growth model for the aggressive behavior factors η_{it}. The linear trend model is described as follows

$$\eta_{it} = \mu_t + \alpha_i + \beta_i \cdot t + \varepsilon_{it}. \tag{40}$$

The variables α_i and β_i are individual level random effects, ε_{it} is the residual of the growth model and μ_t is a time specific parameter. The variance of ε_{it} is also a time specific parameter θ_t. The means of α_i and β_i are fixed to 0 for identification purposes. The intercept and residual variance parameters at time $t = 0$, μ_0 and θ_0, are also fixed to 0 and 1 for identification purposes. The above model can be estimated with the WLSMV estimator and it has 35 parameters only: 9 threshold and loading parameters, 7 factor intercept and factor residual variances as well as the three parameters in the variance covariance matrix of α_i and β_i. The growth model here is a model for the individual growth of the aggressive behavior factor and it reflects the factor variation beyond the average change over time which is modeled with the parameters μ_t. If the linear trend model in (40) does not hold, the β_i random effect will be estimated to 0 and its variance will be estimated to zero as well. An alternative model is the model where the linear change over time includes modeling the change in the factor mean. This can be achieved by removing the parameters μ_t and estimating a mean parameter for β_i. However, often in practical applications the parameters μ_t will not follow a linear trend and such a model would most likely lead to a model misfit. The above model can only be estimated as a multivariate model with the WLSMV estimator. The model is more parsimonious than model (38), however, it relies on the assumption of measurement invariance. The model is also not scalable with respect to time.

8.2 Multilevel SEM

A different approach to modeling time intensive data is two-level modeling where the cluster variable is the individual and the observations at the different time points are the observations within cluster. Consider for example the model

$$P(Y_{pit} = 1) = \Phi(\lambda_p \eta_{it} - \tau_p) \tag{41}$$

$$\eta_{it} = \alpha_i + \beta_i \cdot t + \varepsilon_{it} \tag{42}$$

The variance of ε_{it} is fixed to 1 for identification purposes and the random effects α_i and β_i are individual level growth factors. The random effect α_i

has a mean 0 for identification purposes. The mean of β_i is not zero, i.e., in this model we actually restricted the parameters μ_t from (40) to follow a linear trend. This model is time scalable. The number of parameters is 21: 9 threshold and loading parameters as well as the mean of β_i and the variance parameters of α_i and β_i. This model can be estimated as a two-level model with 3 dimensional numerical integration and it can be estimated with a large number of time points without any computational problems. In fact, the more time points there are the more accurate the estimates of the random effects and the model parameters. Thus this model has the advantage of being time scalable. However, the model has the disadvantage that it assumes measurement invariance.

Another advantage of two-level models for intensive longitudinal data is the fact that we can accommodate more individual level random effects. Consider for example the model

$$P(Y_{pit} = 1) = \Phi(\lambda_p \eta_{it} - \tau_{pi}). \tag{43}$$

$$\eta_{it} = \alpha_i + \varepsilon_{it} \tag{44}$$

$$\tau_{pi} = \tau_p + \epsilon_{pi} \tag{45}$$

The difference between this model and the model (41-42) is that it accommodates individual level variation ϵ_{ip} in the threshold parameters. Such random effects are useful in modeling individual level variation from the factor model. For example, individuals may score lower or higher on a particular item, inconsistent with the rest of the items, due to reasons other than the aggressive behaviors factor. This model has 28 parameters: 9 thresholds, loadings and individual level residual variances as well as the variance parameter of α_i. This model is also scalable in terms of time and it can be estimated with the WLSMV estimator as a two-level model. The model cannot be estimated easily with the ML estimator because such an estimation would require 10 dimensional numerical integration. The Bayes estimator can also be used to estimate this model, in fact, the Bayes estimator can be used to estimate a model that combines all of the features in models (41-42) and (43-45), including the linear trend random effect β_i.

In the above two-level models the latent variable α_i is essentially a between level factor where the between level factor loadings are the same as the within level factor loadings. This however need not be the case. Separate loadings can be estimated for η_{it} and α_i. Other variations of the two-level

models are possible, however, all two-level models assume measurement invariance across time. It is possible to introduce dummy variable predictors for the different time points and thus accommodate threshold non-invariance however such models are not time-scalable due to the many parameters and dummy variables in the model.

8.3 Cross-classified SEM

The cross-classified SEM framework described in this article which accommodates cross-classified modeling as well as random loadings can be used to construct time scalable models that can also accommodate non-invariance for the loadings and threshold parameters. Growth modeling for the factor variable can also be estimated. We use the cross-classified SEM framework because in the intensive longitudinal data the observations are cross-nested within individual and time. Consider the model

$$P(Y_{pit} = 1) = \Phi(\lambda_p \eta_{it} - \tau_{pit}). \tag{46}$$

$$\eta_{it} = \alpha_i + \varepsilon_{it} \tag{47}$$

$$\tau_{pit} = \tau_p + \epsilon_{pi} + \zeta_{pt} \tag{48}$$

This model is very similar to model (43-45). It has an additional random effect ζ_{pt} which has 0 mean and variance ψ_p. This random effect accommodates threshold non-invariance across time. The model has a total of 37 parameters, all of the parameters in model (43-45) plus the 9 random effect variance parameters ψ_p. The next model we consider is the model that accommodates the features of models (41-42), (43-45) and (46-48). This is accomplished by substituting equation (47) with equation (42). This model has just two more parameter: the mean and variance of β_i for a total of 39 parameters. The next modeling extension is to accommodate loading non-invariance across time. Consider the model

$$P(Y_{pit} = 1) = \Phi(\lambda_{pt} \eta_{it} - \tau_{pit}). \tag{49}$$

$$\eta_{it} = \alpha_i + \beta_i \cdot t + \varepsilon_{it} \tag{50}$$

$$\tau_{pit} = \tau_p + \epsilon_{pi} + \zeta_{pt} \tag{51}$$

$$\lambda_{pt} = \lambda_p + \xi_{pt}. \tag{52}$$

The random effects ξ_{pt} allow variation across time in the loading parameters. The mean of ξ_{pt} is 0 and the variance is w_p. This model has just 9 more parameter, the variances parameters w_p for a total of 48 parameters. The final modification we make is to include time specific factor variance. As in the IDFA model the time specific factor variance is introduced by adding a factor model for the random loadings. In addition, we include a time specific factor mean to account for non-invariance of the factor mean that goes beyond the individual growth model. Thus the final model we present is given by the following equations

$$P(Y_{pit} = 1) = \Phi(\lambda_{pt}\eta_{it} - \tau_{pit}). \tag{53}$$

$$\eta_{it} = \nu_t + \alpha_i + \beta_i \cdot t + \varepsilon_{it} \tag{54}$$

$$\tau_{pit} = \tau_p + \epsilon_{pi} + \zeta_{pt} \tag{55}$$

$$\lambda_{pt} = \lambda_p + \lambda_p\sigma_t + \xi_{pt}. \tag{56}$$

The new random effects ν_t and σ_t have zero mean and variances v and σ respectively. The model has a total of 50 parameters. Additional model flexibility can be introduced in the above model by allowing the random loadings to vary not just across time but also across individuals just like the thresholds vary both across time and individuals. Individually specific factor loadings essentially amounts to incorporating the IDFA model into this longitudinal cross-classified framework. For simplicity here we present the parameter estimates for model (53-56). The estimates and standard errors for the item specific parameters are presented in Table 6. The estimates for the variance of the random effects for α_i, β_i, ν_t and σ_t are 1.069(0.076), 0.024(0.003), 0.009(0.020), and 0.008(0.038) respectively. The estimate for the mean of β_i is 0.033(0.019). The Verhagen and Fox (2012) test of significance for the variance components in Table 6 yields that all variance components are significant with the exception of the random effect ζ_{8t}. For this random effect the Bayes factor for the hypothesis $Var(\zeta_{8t}) < 0.001$ is 2.7, i.e., the variance component is only marginally significant. These results imply that intercept and loading non-invariance should be accounted for in the TOCA data. Model (53-56) does so without increasing the number of parameters or without increasing the complexity of the model. This model is easy to interpret because all random effects have a clear purpose. The model is easy to estimate even though it has a total of 31 between level random effects. The model also illustrates the new methodological idea of model customization

Table 6: TOCA cross-classified intensive longitudinal growth model

item	τ_p	Variation across time $Var(\tau_{pit}\|i) = Var(\zeta_{pt})$	Variation across individual $Var(\tau_{pit}\|t) = Var(\epsilon_{pi})$	λ_p	Variation across time $Var(\lambda_{pt}) = Var(\xi_{pt})$
Item 1	-0.781(0.139)	0.064(0.124)	0.153(0.036)	1.017(0.080)	0.013(0.044)
Item 2	-0.690(0.109)	0.021(0.053)	0.171(0.047)	1.413(0.117)	0.011(0.046)
Item 3	1.167(0.300)	0.157(0.520)	0.078(0.041)	1.904(0.207)	0.108(0.328)
Item 4	1.611(0.194)	0.227(0.452)	0.111(0.045)	1.438(0.167)	0.087(0.197)
Item 5	0.045(0.107)	0.029(0.072)	0.231(0.044)	1.319(0.145)	0.058(0.140)
Item 6	1.056(0.211)	0.220(0.338)	0.079(0.037)	1.513(0.130)	0.022(0.079)
Item 7	0.433(0.166)	0.079(0.156)	0.159(0.040)	1.476(0.149)	0.053(0.131)
Item 8	0.279(0.094)	0.006(0.020)	0.120(0.036)	1.324(0.118)	0.024(0.079)
Item 9	-0.374(0.133)	0.043(0.092)	0.200(0.047)	1.369(0.124)	0.025(0.077)

where model parameters can be easily adjusted and customized to a particular subset of the data to achieve better fit without compromising the quality of the model and keeping the model as parsimonious as it can be.

In conclusion, the Cross-classified SEM presented in this section has the advantage that it is time scalable and can accommodate measurement non-invariance. Neither the Longitudinal SEM nor the Multilevel SEM could accomplish both of these tasks at the same time.

9 Conclusion

The Bayesian estimation of structural equation models has become more popular as stable numerical algorithms have been developed. It is now possible to explore models that go beyond the reach of traditional ML and WLS estimators using the Bayesian estimation. Cross classified structural models and random loading models are two such examples. Using these new models it is now possible to address data modeling issues that were not possible to address within the standard structural modeling framework.

References

[1] Asparouhov, T. & Muthén, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. Proceedings of the Joint Statistical Meeting, 2531-2535.

[2] Asparouhov, T. and Muthén, B. (2010a). Bayesian analysis using Mplus: Technical implementation. Technical Report.
http://statmodel.com/download/Bayes3.pdf

[3] Asparouhov T. and Muthén B. (2010b) Bayesian Analysis of Latent Variable Models using Mplus. Technical Report.
http://statmodel.com/download/BayesAdvantages18.pdf

[4] Brennan, R. L. (2001). Generalizability Theory. New York: Springer-Verlag.

[5] Browne, W. J., Goldstein, H. and Rasbash, J. (2001) Multiple membership multiple classification (MMMC) models. Statistical Modelling, 1, 103-124.

[6] Browne, W. J., McCleery, R. H. Sheldon, B. C. and Pettifor, R. A. (2007). Using crossclassified multivariate mixed response models with applications to life history traits in great tits (parus major). Statistical Modelling, 7, 217-238.

[7] Cronbach, L.J., Nageswari, R., & Gleser, G.C. (1963). Theory of generalizability: A liberation of reliability theory. The British Journal of Statistical Psychology, 16, 137-163.

[8] Davidov, E., H. Dulmer, E. Schluter and P. Schmidt (2012) Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. Journal of Cross-Cultural Psychology, 43, 558-575.

[9] De Boeck, P. (2008). Random item IRT models. Psychometrika, 73, 533-559.

[10] De Jong, M.G., Steenkamp, J.-B.E.M. (2009) Finite Mixture Multilevel Multidimensional Ordinal IRT Models for Large Scale Cross-Cultural Research, Psychometrika, 75, 3-32.

[11] De Jong, M.G., Steenkamp, J.-B.E.M., & Fox, J.-P. (2007). Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. Journal of Consumer Research, 34, 260-278.

[12] Fielding, A., & Goldstein, H. (2006). Cross-classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review. Research Report RR791. Birmingham, UK: Department for Education and Skills, University of Birmingham.

[13] Goldstein, H. (2011). Multilevel Statistical Models, 4rd Edition. London, Arnold.

[14] Gonzalez, De Boeck, Tuerlinckx (2008) A Double-Structure Structural Equation Model for Three-Mode Data. Psychological Methods, 337 - 353.

[15] Fox, J.P. (2010) Bayesian Item Response Theory. Springer.

[16] Luo, W., & Kwok, O-M. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. Multivariate Behavioral Research, 44, 182-212.

[17] Luo, W., & Kwok, O.-M. (2012). The Consequences of Ignoring Individuals' Mobility in Multilevel Growth Models: A Monte Carlo Study. Journal of Educational and Behaviorial Statistics, 37, 31-56.

[18] Ialongo, N. S., Werthamer, L., Kellam, S. G., Brown, C. H., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression and antisocial behavior. American Journal of Community Psychology, 27, 599-642.

[19] Jahng S., Wood, P. K.,& Trull, T. J., (2008) Analysis of Affective Instability in Ecological Momentary Assessment: Indices Using Successive Difference and Group Comparison via Multilevel Modeling. Psychological Methods, 13, 354-375.

[20] Marcoulides, G. A. (1999). Generalizability theory: Picking up where the Rasch IRT model leaves off? In S.E. Embretson & S.L. Hershberger et al. (Eds.). The new rules of measurement: What every psychologist

and educator should know. (pp. 129-152). Mahwah, NJ, USA: Lawrence Erlbaum.

[21] Marsh, H. W. and Hocevar, D. (1991) Multidimensional perspective on students evaluations of teaching effectiveness: the generality of factor structures across academic discipline, instructor level, and course level, Teaching and Teacher Education, 7, 9-18.

[22] Meng, X. (1994). Posterior Predictive p-Values. Annals of Statististic 22, 1142-1160.

[23] Muthén, B. and Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. Psychological Methods, 17, 313-335.

[24] Muthén, B. O. (1994). Multilevel covariance structure analysis. Sociological Methods & Research 22(3), 376-398.

[25] Muthén, L.K. and Muthén, B.O. (1998-2014). Mplus Users Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén

[26] Rasbash, J. and Goldstein, H. (1994) Efficient analysis of mixed hierarchical and crossed random structures using a multilevel model. Journal of Behavioural Statistics 19, 337-350.

[27] Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross sectional and longitudinal research. Journal of Educational Statistics 18: 321-349.

[28] Raudenbush, S.W. and Bryk, A.S. (2002). Hierarchical Linear Models. Thousand Oaks: Sage Publications.

[29] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B, 64, 583639.

[30] Verhagen, A. J. & Fox, J.-P. (2012) Bayesian Tests of Measurement Invariance. British Journal of Mathematical and Statistical Psychology.

[31] Walls, T.A. & Schafer, J.L. (2006). Models for Intensive Longitudinal Data. New York : Oxford University Press.