Latent variable interactions using maximum-likelihood and Bayesian estimation for single- and two-level models

Tihomir Asparouhov & Bengt Muthén

Mplus

www.statmodel.com

Mplus Web Notes: No. 23

May 3, 2019

1 Introduction

This web note discusses analysis of structural equation models with latent variable interactions. With a focus on maximum-likelihood estimation, Section 2 discusses interpretation, model testing, explained variance, standardization, and plotting of effects for models with latent variable interactions. Section 3 discusses Bayesian estimation and its advantages over maximum-likelihood estimation, particularly for twolevel models with moderation. The Appendix presents standardization in matrix terms.

2 Latent variable interactions using ML

Structural equation modeling with latent variable interactions has been discussed with respect to maximum-likelihood estimation in Klein and Moosbrugger (2000). Multivariate normality is assumed for the latent variables. The ML computations are heavier than for models without latent variable interactions because numerical integration is needed. For an overview of the ML approach and various estimators suggested in earlier work, see Marsh et al. (2004). Arminger and Muthén (1998), Klein and Muthén (2007), Cudeck et al. (2009), and Mooijaart and Bentler (2010) discuss alternative estimators and algorithms.

2.1 Model interpretation

As an example, consider the latent variable interaction model of Figure 1. The figure specifies that the factor η_3 is regressed on η_1 and η_2 as well as the interaction between η_1 and η_2, as shown by the structural equation

$$\eta_3 = \beta_1\, \eta_1 + \beta_2\, \eta_2 + \beta_3\, \eta_1 \times \eta_2 + \zeta_3. \tag{1}$$

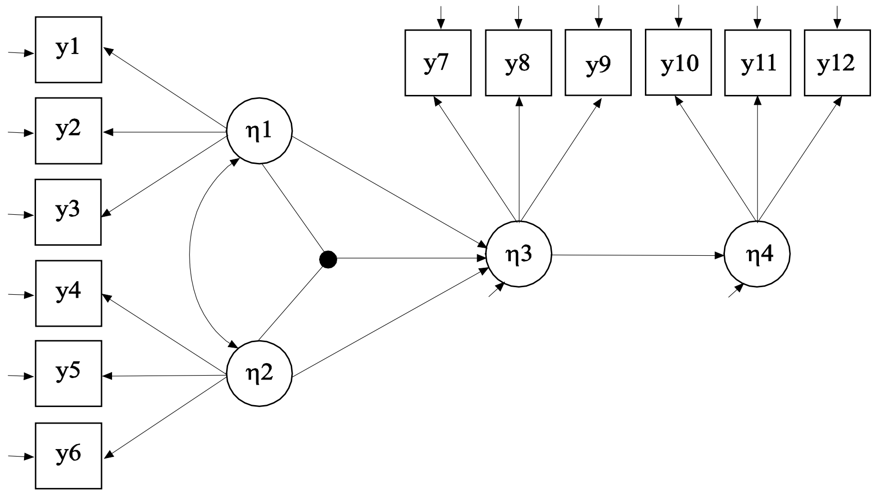
The interaction variable $\eta_1 \times \eta_2$ involves only one parameter, the slope β_3. The interaction variable does not have a mean or a variance parameter. It does not have parameters for covariances with other variables. It can also not be a dependent variable. As is seen in Figure 1, the model also contains a second structural equation where η_4 is linearly regressed on η_3, so that there is no direct effect on η_4 from η_1 and η_2, or their interaction.

For ease of interpretation the (1) regression can be re-written in the equivalent form

$$\eta_3 = (\beta_1 + \beta_3\, \eta_2)\, \eta_1 + \beta_2\, \eta_2 + \zeta_3, \tag{2}$$

where $(\beta_1 + \beta_3\, \eta_2)$ is a moderator function (Klein & Moosbrugger, 2000) so that the β_1 strength of influence of η_1 on η_3 is moderated by $\beta_3\, \eta_2$. The choice of moderator when translating (1) to (2) is arbitrary from an algebraic point of view, and is purely a choice based on ease of substantive interpretation. As an example, Cudeck et al. (2009) considers school achievement (η_3) influenced by general reasoning (η_1), quantitative ability (η_2), and their interaction. In line with (2) the interaction is expressed as quantitative ability moderating the influence of general reasoning on school achievement. Plotting of interactions further aids the interpretation as discussed in Section 2.5.

Figure 1: Structural equation model with interaction between latent variables



2.2 Model testing

As pointed out in Mooijaart and Satorra (2009), for some SEM models, the likelihood-ratio χ^2 obtained by ML for models without latent variable interactions is not sensitive to incorrectly leaving out latent variable interactions. For example, the model of Figure 1 without the interaction term $\beta_3\ \eta_1 \times \eta_2$ fits data generated as in (1) perfectly. This is due to general maximum-likelihood results on robustness to non-normality (Satorra, 1990, 2002). Misfit for that model can be detected only by considering higher-order moments than the second-order variances and covariances of the outcomes. For other SEM models, omitted interaction terms can be detected by the chi-square test of fit, see Section 3.2 below. In such cases, traditional model modification guidance based on the chi-square test of fit would be incorrect and would lead to linear models that fit the first and second order moments approximately but would fail to discover the need for interaction terms. All this suggests that the standard chi-square test of fit has fairly limited capabilities when dealing with interaction modeling.

Without involving higher-order moments, a reasonable modeling strategy is to first fit a model without interactions and then add an interaction term. The significance of the interaction can be tested by either a z-test or a likelihood-ratio χ^2 difference test. Likelihood-ratio or Wald tests can be used to test the joint significance of several interaction terms.

2.3 Mean, variance, and R^2

To compute a dependent variable mean, variance, and R^2 for models with latent variable interactions, the following results are needed. The covariance between two variables x_j and x_k is defined as

$$Cov(x_j, x_k) = E(x_j\ x_k) - E(x_j)\ E(x_k), \tag{3}$$

so that the variance is obtained as

$$Cov(x_j, x_j) = V(x_j) = E(x_j^2) - [E(x_j)]^2. \tag{4}$$

With $E(x_j) = 0$ or $E(x_k) = 0$, (3) gives the mean of a product

$$E(x_j\ x_k) = Cov(x_j, x_k). \tag{5}$$

Assuming multivariate normality for four random variables x_i, x_j, x_k, x_l any third-order moment about the mean (μ) is zero (see, e.g., Anderson, 1984),

$$E(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k) = 0, \tag{6}$$

while the fourth-order moment about the mean is a function of covariances,

$$E(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l) = \sigma_{ij}\,\sigma_{kl} + \sigma_{ik}\,\sigma_{jl} + \sigma_{il}\,\sigma_{jk}, \tag{7}$$

where for example $\sigma_{jk} = Cov(x_j, x_k)$ and $\sigma_{kk} = Var(x_k)$. This gives

$$E(x_j\,x_k\,x_j\,x_k) = V(x_j)\,V(x_k) + 2\,[Cov(x_j, x_k)]^2, \tag{8}$$

so that the variance of a product is obtained as

$$V(x_j\,x_k) = E(x_j\,x_k\,x_j\,x_k) - [E(x_j\,x_k)]^2 \tag{9}$$

$$= V(x_j)\,V(x_k) + 2\,[Cov(x_j, x_k)]^2 - [Cov(x_j, x_k]^2 \tag{10}$$

$$= V(x_j)\,V(x_k) + [Cov(x_j, x_k)]^2, \tag{11}$$

Consider the application of these results to the mean and variance of the factor η_3 in (1) of Figure 1. Because of zero factor means, using (5) the mean of η_3 in (1) is obtained as

$$E(\eta_3) = \beta_1\,0 + \beta_2\,0 + \beta_3\,E(\eta_1\,\eta_2) + 0 \tag{12}$$

$$= \beta_3\,Cov(\eta_1, \eta_2). \tag{13}$$

Using (4), the variance of η_3 is

$$V(\eta_3) = E(\eta_3\,\eta_3) - [E(\eta_3)]^2, \tag{14}$$

where the second term has already been determined. As for the first term, multiplying the right-hand side of (1) by itself results in products of two, three, and four factors. Expectations for three- and four-factor terms are simplified by the following two results, assuming bivariate normality and zero means for η_1 and η_2. All third-order moments $E(\eta_i\ \eta_j\ \eta_k)$ are zero by (6). The formula (8) is used to obtain the result

$$E(\eta_1\ \eta_2\ \eta_1\ \eta_2) = V(\eta_1)\ V(\eta_2) + 2\ [Cov(\eta_1, \eta_2)]^2. \tag{15}$$

Collecting terms, it follows that the variance of η_3 is obtained as

$$V(\eta_3) = \beta_1^2\ V(\eta_1) + \beta_2^2\ V(\eta_2) + 2\ \beta_1\ \beta_2\ Cov(\eta_1, \eta_2) + \beta_3^2\ V(\eta_1\ \eta_2) + V(\zeta_3), \tag{16}$$

where by (11)

$$V(\eta_1\ \eta_2) = V(\eta_1)\ V(\eta_2) + [Cov(\eta_1, \eta_2)]^2, \tag{17}$$

R-square for η_3 can be expressed as usual as

$$[V(\eta_3) - V(\zeta_3)]/V(\eta_3). \tag{18}$$

Using (16), the proportion of $V(\eta_3)$ contributed by the interaction term can be quantified as (cf. Mooijaart & Satorra, 2009; p. 445)

$$\beta_3^2\ [V(\eta_1)\ V(\eta_2) + [Cov(\eta_1, \eta_2)]^2]/V(\eta_3). \tag{19}$$

Consider as a hypothetical example the latent variable interaction model of

Figure 2: Structural equation model with interaction between an exogenous and an endogenous latent variable

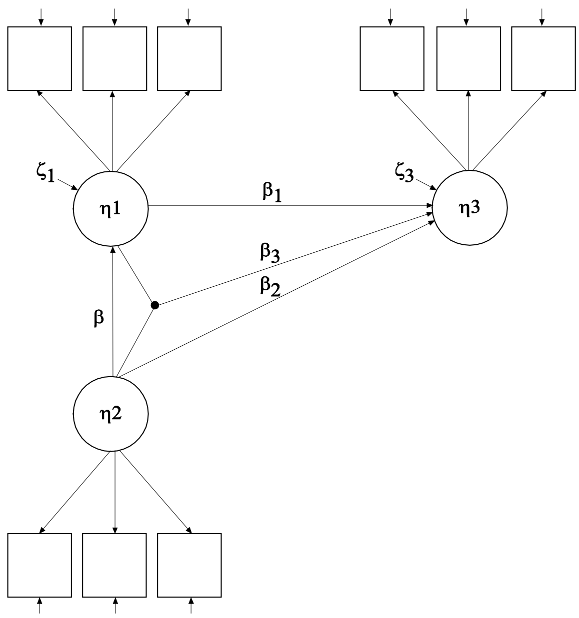


Figure 2. Here, the latent variable interaction is between an exogenous and an endogenous latent variable. This example is useful to study the details of how to portray the model. The structural equations are

$$\eta_1 = \beta\ \eta_2 + \zeta_1, \tag{20}$$

$$\eta_3 = \beta_1\ \eta_1 + \beta_2\ \eta_2 + \beta_3\ \eta_1 \times \eta_2 + \zeta_3. \tag{21}$$

Let $\beta = 1$, $\beta_1 = 0.5$, $\beta_2 = 0.7$, $\beta_3 = 0.4$, $V(\eta_2) = 1$, $V(\zeta_1) = 1$, and $V(\zeta_3) = 1$. This implies that $V(\eta_1) = \beta^2\ V(\eta_2) + V(\zeta_1) = 1^2 \times 1 + 1 = 2$ and $Cov(\eta_1, \eta_2) = \beta\ V(\eta_2) = 1 \times 1 = 1$. Using (16), $V(\eta_3) = 3.17$. The η_3 R-square is 0.68 and the variance percentage due to the interaction is 15%.

2.4 Standardization

Because latent variables have arbitrary metrics, it is useful to also present interaction effects in terms of standardized latent variables. Noting that (21) is identical to (1), the model interpretation is aided by considering the moderator function $(\beta_1 + \beta_3\,\eta_2)\,\eta_1$ of (2), so that η_2 moderates the η_1 influence on η_3.

As usual, standardization is obtained by dividing by the standard deviation of the dependent variable and multiplying by the standard deviation of the independent variable. The standardized β_1 and β_3 coefficients in the term $(\beta_1 + \beta_3\,\eta_2)\,\eta_1$ are obtained by dividing both by $\sqrt{V(\eta_3)} = \sqrt{3.17}$, multiplying β_1 by $\sqrt{V(\eta_1)} = \sqrt{2}$, and multiplying β_3 by $\sqrt{V(\eta_1)}\,\sqrt{V(\eta_2)} = \sqrt{2}$. This gives a standardized $\beta_1 = 0.397$ and a standardized $\beta_3 = 0.318$. The standardization of β_3 is in line with Wen, Marsh, and Hau (2010; equation 10). These authors discuss why standardization of β_3 using $\sqrt{V(\eta_1)}\,\sqrt{V(\eta_2)}$ is preferred over using $\sqrt{V(\eta_1 \times \eta_2)}$.

The standard deviation change in η_3 as a function of a one standard deviation change in η_1 can now be evaluated at different values of η_2 using the moderator function. At the zero mean of η_2, a standard deviation increase in η_1 leads to a 0.397 standard deviation increase in η_3. At one standard deviation above the mean of η_2, a standard deviation increase in η_1 leads to a $0.397 + 0.318 \times 1 = 0.715$ standard deviation increase in η_3. At one standard deviation below the mean of η_2, a standard deviation increase in η_1 leads to a $0.397 - 0.318 \times 1 = 0.079$ standard deviation increase in η_3. In other words, the biggest effect of η_1 on η_3 occurs for subjects with high values on η_2.

A more general treatment of standardization in matrix terms is given in the

Appendix.

2.5 Plotting of interactions

The interaction can be plotted as in Figure 3. Using asterisks to denote standardization, consider the rearranged (21),

$$\eta_3^* = (\beta_1^* + \beta_3^* \, \eta_2^*) \, \eta_1^* + \beta_2^* \, \eta_2^* + \zeta_3^*. \tag{22}$$

Using (22), the three lines in the figure are expressed as follows in terms of the conditional expectation function for η_3^* at the three levels of η_2^*,

$$E(\eta_3^* | \eta_1^*, \eta_2^* = 0) = \beta_1^* \, \eta_1^*, \tag{23}$$

$$E(\eta_3^* | \eta_1^*, \eta_2^* = 1) = (\beta_1^* + \beta_3^*) \, \eta_1^* + \beta_2^*, \tag{24}$$

$$E(\eta_3^* | \eta_1^*, \eta_2^* = -1) = (\beta_1^* - \beta_3^*) \, \eta_1^* - \beta_2^*. \tag{25}$$

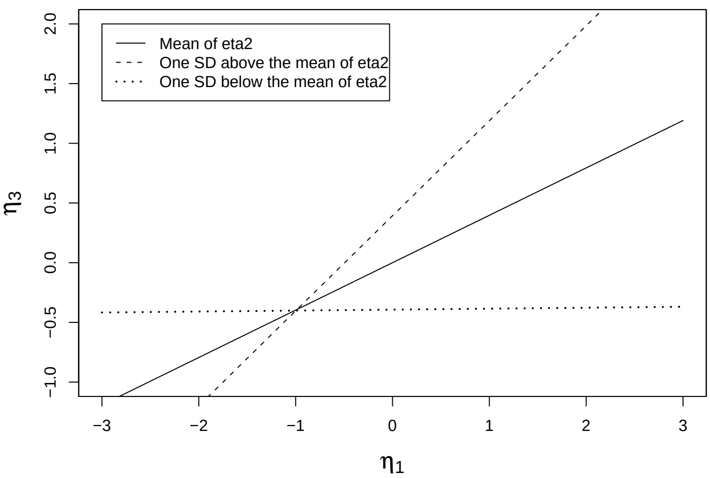
$$\tag{26}$$

Here, the standardized value $\beta_2^* = \beta_2 \times \sqrt{V(\eta_2)} / \sqrt{V(\eta_3)} = 0.7 \times 1 / \sqrt{3.17} = 0.393$.

3 Bayesian estimation

The interactions between a latent variable and an observed variable can be estimated with the maximum likelihood estimator using a closed form expression for the likelihood, see Muthén and Asparouhov (2003). Because numerical integration is not used in that case, the estimation is efficient and can accommodate any number of interaction terms. Interactions between two latent variables,

Figure 3: Interaction plot for structural equation model with interaction between an exogenous and an endogenous latent variable



however, do not lead to a closed form expressions for the likelihood and cannot be estimated with the maximum likelihood method without numerical integration, see Klein & Moosbrugger (2000). In Mplus, the dimension of numerical integration corresponds to the number of latent variables used in the interaction terms. Numerical integration with more than 3 dimensions of integration, i.e., with more than 3 latent variables is generally computationally intractable. It is possible to estimate models with larger number of dimensions of integration using montecarlo integration in Mplus or using quadrature integration with very few integration points per dimension, however, such estimation often lacks precision and results in non-convergence.

By using the Bayesian estimation, however, we can resolve these limitations of the ML estimator, and estimate models with any number of latent variables and interaction terms. Consider the following general SEM model with interactions. Let Y_p denote the observed dependent variables, $p = 1, ..., P$, η_m denote the latent variables in the model, $m = 1, ..., M$ and X_q denote the covariates, $q = 1, ...Q$. The general interaction model can be described as follows

$$Y_p = \nu_p + \sum_{i=1}^{M} \lambda_{pi}\eta_i + \sum_{i=1}^{M}\sum_{j=i}^{M} \gamma_{pij}\eta_i\eta_j + \varepsilon_p \tag{27}$$

$$\eta_m = \alpha_m + \sum_{i=1}^{M} \beta_{mi}\eta_j + \sum_{i=1}^{M}\sum_{j=i}^{M} \delta_{mij}\eta_i\eta_j + \sum_{q=1}^{Q} \kappa_{mq}X_q + \xi_m \tag{28}$$

where ν_p, λ_{pi}, γ_{pij}, α_m, β_{mi}, δ_{mij}, κ_{mq} are model parameters and ξ_m and ε_p are normally distributed residuals.

To estimate this model with the Bayesian method we follow the MCMC estimation framework described in Asparouhov and Muthén (2010a) for the estimation of the general SEM model. All aspects of that estimation method remain the same with the exception of one - the Gibbs sampler step for generating the latent variables η_m. In the standard SEM model the posterior distribution for the latent variables used in the Gibbs sampler is a multivariate normal distribution. Due to the interaction terms, however, the posterior distribution will not be normal for the above model. To resolve this issue, we split the Gibbs sampler for the latent variables so that each latent variable is generated conditional on all other latent variables, i.e., we replace the Gibbs sampler that generates $\eta_1, \eta_2, ..., \eta_M$ simultaneously with a Gibbs sampler with m steps that generates one latent variable at a time using the posterior distributions

$$[\eta_1 | \eta_2, \eta_3, ..., \eta_M, *] \tag{29}$$

$$[\eta_2|\eta_1, \eta_3, ..., \eta_M, *] \tag{30}$$

....

$$[\eta_M|\eta_1, \eta_2, ..., \eta_{M-1}, *] \tag{31}$$

The advantage of this approach is that the above univariate distributions are easier to solve for. There are two separate cases. Consider the posterior distribution $[\eta_1|\eta_2, \eta_3, ..., \eta_M, *]$. The first and the simpler case is the situation when $\gamma_{p11} = \delta_{m11} = 0$ for each p and m. In that case, the quadratic term η_1^2 is not included in the model. The variable η_1 can be included in interactions terms such as $\eta_1\eta_2$, $\eta_1\eta_3$, but because the variables η_2, η_3, ... are conditioned on, the model remains linear in terms of η_1. Therefore in this case the posterior $[\eta_1|\eta_2, \eta_3, ..., \eta_M, *]$ is the normal distribution with closed form expression that can be computed as in Asparouhov and Muthén (2010a).

The second case is the situation when η_1^2 is included in the model, i.e., some of the parameters γ_{p11} or δ_{m11} are not zero. In that case the posterior distribution of η_1 is not explicit and does not have a closed form expression. We utilize the Metropolis-Hastings algorithm. To do that we specify the jumping distribution J as follows. If the current value for η_1 is η_1^* we consider the model where the interaction terms η_1^2 are replaced by $\eta_1\eta_1^*$. The posterior distribution of η_1 from that new model is a normal distribution and has a closed form expression which we choose as the jumping distribution J for drawing a proposal value for η_1, i.e., we draw a new value η_1^{**} from this distribution. The new value is accepted with probability $min(1, A)$ where A is

$$A = \frac{p(y, \eta_{-1}, \eta_1^{**}|x)J(\eta_1^*|\eta_1^{**})}{p(y, \eta_{-1}, \eta_1^*|x)J(\eta_1^{**}|\eta_1^*)} \tag{32}$$

where p is the normal densities for that conditional distribution which can be derived from the model for ε_p and ξ_m in a sequential way under general regularity conditions (ex. η_1 is not regressed on η_1^2 or more generally - no reciprocal interactions). The vector η_{-1} denotes all η variables except the first.

The above estimation method easily extends to two-level models and models with categorical dependent variables as it is done in Asparouhov and Muthén (2010a). In the two-level model the dependent variables are split as within and between and each of the two parts follow the SEM model with interactions (27-28), i.e., the within portion of a variable can be predicted by interactions of within level latent variables, while the between portion of the variable can be predicted by interactions of between level latent variables. Cross-interactions of within level latent variables and between level latent variables are easily accommodated as well since such terms are essentially random slope coefficients for within level latent variables.

The interaction model can also be extended to incorporate interactions between a latent variable and an observed variable. In Mplus this can be done directly by specifying the interaction effect using the XWITH option between a latent variable and an observed variable. It can also be done by introducing a latent variable "behind" the observed variable (i.e. the observed variable is a perfect indicator for the latent variable) and then using the XWITH option for the two latent variables. The second approach is less efficient as it generally yields MCMC chains with worse mixing quality, however, if the observed variable has missing values it would be the only available approach in Mplus. Another condition that requires the second approach is the situation where the observed variables is in a two-level model and is latent-centered.

The Bayesian estimation described here is fairly close to the one described in Arminger and Muthén (1998). There are two main differences. The first one is that the latent variables are updated one at a time which allows us to use conjugate posterior distributions most cases instead of the less efficient Metropolis-Hastings algorithm. The second difference is in the proposal distribution used in the Metropolis-Hastings algorithm. The proposal distribution used in Arminger and Muthén (1998) is the same across individuals and is based on the model estimated distribution, i.e., it ignores the measurement part of the model. The proposal distribution used in this algorithm incorporates the measurement part of the model, including the quadratic terms where the latent variables are involved. This makes the proposal distribution subject specific and very close to the desired posterior distribution, which leads to a well mixing MCMC estimation.

In the next sections we illustrate the methodology with several examples.

3.1 Factor analysis with interactions

In this example we consider a factor analysis model where 5 factors are measured by 3 indicator variables each, i.e., we have a total of 15 observed variables. All loadings are set to 1, intercepts are set to 0, residual and factor variances are set to 1, factor correlations are set to 0.3. We add the following three interaction terms in the model $\eta_1\eta_2$, $\eta_1\eta_5$ and $\eta_3\eta_4$. The effects of these interaction terms on the observed variables are all zero except for the following 3 effects $\eta_1\eta_2$ on Y_1 is set to $\gamma_{112} = -0.25$, the effect of $\eta_1\eta_5$ on Y_1 is set to $\gamma_{115} = 0.25$ and the effect of $\eta_3\eta_4$ on Y_7 is set to $\gamma_{734} = 0.25$. The model is described by the following equations

$$Y_1 = \nu_1 + \lambda_{11}\eta_1 + \gamma_{112}\eta_1\eta_2 + \gamma_{115}\eta_1\eta_2 + \varepsilon_1 \tag{33}$$

$$Y_p = \nu_p + \lambda_{p1}\eta_1 + \varepsilon_p, p = 2, 3 \tag{34}$$

$$Y_p = \nu_p + \lambda_{p2}\eta_2 + \varepsilon_p, p = 4, 5, 6 \tag{35}$$

$$Y_7 = \nu_7 + \lambda_{73}\eta_3 + \gamma_{734}\eta_3\eta_4 + \varepsilon_7 \tag{36}$$

$$Y_p = \nu_p + \lambda_{p3}\eta_3 + \varepsilon_p, p = 8, 9 \tag{37}$$

$$Y_p = \nu_p + \lambda_{p4}\eta_4 + \varepsilon_p, p = 10, 11, 12 \tag{38}$$

$$Y_p = \nu_p + \lambda_{p5}\eta_5 + \varepsilon_p, p = 13, 14, 15 \tag{39}$$

We compare the Bayesian estimation method and the ML - montecarlo integration method where the number of integration points is set to 500 and 1000. Using 100 data sets of size 1000 we estimate the correct interaction model with the two estimators and report the results in Table 1 for the interaction effects. All three estimation methods yield acceptable results, however, the ML method with 500 integration points shows significantly larger MSE for the estimates. The reduced precision in the log-likelihood computation yields reduced precision in the ML estimates. The convergence rates for the Bayes method and the ML(1000) is 100% while the convergence for the ML(500) is 98%, i.e., a slight drop in the convergence rates. The computational time for Bayes and ML(500) is approximately the same while the computational time for ML(1000) is about twice that. Usually the computational time with montecarlo integration is proportional to the number of integration points. Overall the conclusion from this simulation is that the Bayes estimator appears to be the best in terms of computational time and precision of the estimates, however, the differences with the ML estimator are not large. Increasing the number of integration points increases the precision of the estimates, although, it is not a priori clear how to determine the optimal number of integration points. In the above simulation, increasing the number of integration points to 5000 did not improve the precision of the estimates in terms of MSE but increased substantially the computation time per replication. This is a clear advantage of the Bayesian estimation as it removes the uncertainty of the number of integration points. Essentially, it automatically determines the amount of computation that has to be done to obtain precise estimates.

We also considered estimating the modified model where all indicator variables are regressed on the three interaction terms. In a typical application that would be the most likely scenario. The effect of estimating this model with many more parameters on the convergence rate is that the convergence rate for ML(500) dropped to 95% while for the Bayes estimator it remained at 100%. This is a slightly bigger drop on the convergence rate and most likely bigger convergence problems should be expected for models that have flatter likelihoods where estimation precision is more important. Nevertheless we can see here that the convergence rates remain high. This is primarily due to the fact that there are only 5 dimensions of integration. In the next example, where we consider a similar two-level model, the situation is completely different and the montecarlo integration method shows no such promise, i.e., the Bayes method appears to be the only alternative.

Consider the following two-level factor analysis model with interactions. The within level model would be identical to the model we considered above, while the between level model simply consists of random intercepts with variance 1. We generate and analyze 100 data sets which consist of 100 clusters of size 20. The ML method as implemented in Mplus uses 20 dimensions of integration,

Table 1: Factor analysis with interactions: absolute bias/coverage/MSE

Parameter	True Value	Bayes	ML(500)	ML(1000)
γ_{112}	-.25	.01/.99/.002	.01/.95/.006	.00/.99/.002
γ_{115}	.25	.01/.95/.002	.01/.96/.004	.00/.96/.002
γ_{734}	.25	.00/.98/.002	.02/.96/.003	.00/.99/.002

Table 2: Factor analysis with interactions: absolute bias/coverage/MSE

Parameter	True Value	Bayes
γ_{112}	-.25	.010/.95/.002
γ_{115}	.25	.00/.90/.002
γ_{734}	.25	.00/.95/.001

5 on the within level and 15 on the between level for each of the 15 random intercept variables, one for each observed variable. Using the ML method with 5000 integration points the convergence rate we obtained is 0%. On the other hand, the Bayes method has 100% convergence. The results for the interaction parameters are presented in Table 2. The Bayes estimation is clearly the only alternative for this model and the method performs well. The estimation time for this model using 4 processors is approximately 15 seconds per iteration. The ML montecarlo method with 500 integration points took approximately 20 minutes per iteration (while no convergence was actually achieved).

One important aspect of interaction modeling is the question regarding which interaction effects should be considered for model inclusion in the absence of any substantive guidance. In our example there are a total of 225 possible interaction parameters γ_{pij}. These parameters would naturally be in addition to any cross-loading parameters. The total number of parameters can easily become quite large. In such a case, one can include all these parameters within the BSEM framework, see Muthén and Asparouhov (2012), where these additional possible parameters will be included with tiny priors centered at zero. In that exploratory framework, interaction effects that should be included will "escape" the tiny prior to exhibit significance while at the same time allowing the main factor model to be adjusted accordingly for the effect of the included interaction terms.

Another approach that can be utilized for exploratory purposes is to generate plausible values, see Asparouhov and Muthén (2010b), for the factor analysis model without the interactions. As a second step then compute the residuals ε_p and all interaction terms $\eta_i\eta_j$ using these plausible values. As a third step one can compute the sample correlation matrix (using Mplus type=imputation option) for all of these variables and select for model inclusion the interactions that have substantial correlations with the residual variables ε_p.

3.2 The effect of ignoring interaction terms

An important question that should be addressed here is why we need to incorporate interaction effects in the SEM models. Perhaps ignoring the interaction effects would lead to no essential problems for the factor analysis. In principle the factor analysis model is estimated from the first and the second order sample statistics, while interaction terms tend to be needed to fit higher order moments. It is conceivable from that point of view that ignoring interaction terms might have no effect on the SEM model. This, however, is not the case and we will illustrate this point with several CFA and EFA simulation studies.

We generate data using a model similar to the model (33-39) used in the previous section with some small modifications. In this section we use the same interaction terms as in (33-39) but we include two additional non-zero interaction effects, i.e., a total of 5 non-zero interaction effects: $\delta_{112} = -.5$, $\delta_{115} = .5$, $\delta_{212} = .5$, $\delta_{412} = .5$, $\delta_{734} = .5$. We generate 100 data sets of size 1000 and we analyze the data using the CFA model without the interaction terms. We utilize the ML and the MLR estimators in Mplus. The MLR estimator is generally expected to perform better given that the interaction terms would be incorporated in the residuals of the CFA model, i.e., are expected to have non-normal distributions, which is where the advantage of the MLR estimator is.

The results of the simulation show that there is little difference between ML and MLR chi-square statistics and both reject the model 95% of the time. On the other hand all approximate fit indices accept the model: the average value for the RMSEA is 0.02, the average value for the SRMR is 0.02, the average value for the CFI is 0.99, and the average value for the TLI is 0.98. One can conclude from this example that approximately fitting models, rejected by the exact chi-square test of fit, may indeed be models that have omitted interaction terms (among other types of minor misspecifications). Table 3 contains the results for the second and third factor loadings (the first factor loading is fixed to 1) for the ML and MLR estimators with omitted interaction terms and the Bayes estimator with the interaction terms included. First we note that the Bayes estimator yields low coverage for the first loading even though the bias is negligible. Usually such situations can be resolved by running a longer MCMC sequence, instead of relying on the default convergence settings. The ML run shows bias for both loadings but particularly large bias for the the third loading and substantial drop in coverage. Using the MLR estimator improves the coverage but not sufficiently. We conclude

Table 3: Factor analysis with omitted interactions: absolute bias(coverage)

Parameter	True Value	Bayes	ML	MLR
λ_{21}	1	.02(.81)	.04(.88)	.04(.93)
λ_{31}	1	.01(.92)	.15(.60)	.15(.67)

here that omitted interaction terms can change the factor structure and bias the factor loadings. This occurs even when the factors are uncorrelated. The model estimation with the omitted interaction terms will attempt to incorporate the interaction terms implied covariance within one of the existing factors which in turn will distort the measurement model for that factor.

Mooijaart and Satorra (2009) point out that for some SEM models the likelihood-ratio test cannot detect omitted interaction terms. As the above example shows, however, this does not apply to all models. Special models where interaction terms between latent variables do not affect directly the observed variables but only other latent variables can be expected to produce correct chi-square even when the interaction terms are ignored.

Now we turn our attention to the effect of omitted interaction terms on EFA. We analyze the same data as above with a 5 factor EFA model and a 6 factor EFA model. Using the chi-square test of fit, we reject the 5 factor model 93% of the time and reject the 6 factor model 9% of the time, i.e., 84% of the time we conclude that the number of factors is 6. Thus omitted interactions can lead to incorrect number of factors in EFA. In addition, Table 4 shows the results for several factor loading estimates for the 5-factor EFA model. Here the factor loadings are biased as well. This fact has some implications regarding the question of how to include the interactions within the EFA estimation. One possible approach is to use the

Table 4: EFA with omitted interactions: absolute bias(coverage)

Parameter	True Value	MLR
λ_{11}	1	.11(.64)
λ_{21}	1	.12(.59)
λ_{31}	1	.11(.64)
λ_{22}	0	.12(.63)

ESEM-within-CFA approach described in Marsh et al. (2013). Because of the biases shown in Table 4, however, such an approach may still result in biased loading structure even after including the interaction terms. Further adjustments might be necessary in such situations. The results in Table 4, see λ_{22}, also show that small cross-loadings can appear in the model due to the omitted interaction terms. It is also worth noting here that the 6-factor EFA model, picked by the chi-square test of fit, has and additional (sixth) factor of somewhat uninterpretable quality. This factor has multiple medium range loadings with large standard errors that appear to be statistically insignificant. This kind of phenomenon also appears quite often in real data EFA, i.e., it could potentially be due to omitted interaction terms.

One issue that may appear as a stumbling block for the Bayesian interaction modeling is the lack of fit statistics. Neither DIC nor PPP are available in Mplus at this time. It is possible however to make informed model modifications in the interaction framework by evaluating the significance of the interaction coefficients, i.e., if an interaction term has a significant effect as established by the credibility interval it should be included in the model and if the effect is insignificant the interaction terms is not needed and can be removed from the model.

3.3 Two-level moderation analysis

Preacher et al. (2016) describe several two-level moderation models with interactions among predictors at the within level, the between level and across the two levels. The authors used Mplus to estimate the models via maximum-likelihood with numerical integration. With the release of Mplus 8.3 these models can now be estimated with the Bayesian method. In this section we compare the accuracy, speed and robustness between the different estimation methods using simulation studies. The scripts we use for the simulation studies are taken directly from the Supplemental materials of Preacher et al. (2016), although in certain cases we have simplified the inputs. Such simplifications, however, do not alter the models. We also preserve the notation used in that article for quick reference. For example, model A1 in Preacher et al. (2016) refers to the interaction model [Within part of L1 moderator] x [Within part of L1 predictor]. In the next 8 sections we present simulation studies on the first 8 examples in the supplemental materials in Preacher et al. (2016) and we preserve the order of these examples.

We illustrate below that the use of the Bayesian method allows us to more fully pursue these moderation models. The Bayesian estimation of these models is faster, simpler, and more robust (more likely to converge) than the maximum-likelihood estimation.

3.3.1 Model A1: [Within part of L1 moderator] x [Within part of L1 predictor]

Suppose that Y_{ij}, X_{ij} and Z_{ij} are the observed variables for individual i in cluster j. The model can be described by the following equations. First we decompose the variables X_{ij} and Z_{ij} as within-between

$$X_{ij} = X_i + X_{.j} \tag{40}$$

$$Z_{ij} = Z_i + Z_{.j} \tag{41}$$

where $X_{.j}$ and $Z_{.j}$ are the between level parts of the variables (i.e. their cluster specific means), which are assumed to be normally distributed latent variables. The moderation model is then given as follows

$$Y_{ij} = \beta_{0j} + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + \varepsilon_{ij} \tag{42}$$

where

$$\beta_{0j} = \gamma_{00} + \gamma_{01} X_{.j} + \gamma_{02} Z_{.j} + u_{0j} \tag{43}$$

and ε_{ij} and u_{0j} are normally distributed zero mean residuals. To generate the data we use the model parameters in Preacher et al. (2016) supplemental materials. We generate 100 data sets with 100 clusters of size 10. Using the Bayes method, the estimation converged in all 100 replications and the estimation took only a couple of seconds per replication. Using maximum likelihood with quadrature integration as in Preacher et al. (2016) we obtained 100% non-convergence. This simulation is different from the one used in the original article because it is based on smaller number of smaller clusters, and indeed for larger samples convergence can be achieved. Using Montecarlo integration (MLMC) for this estimation did not yield any convergence either. Table 5 contains the results of the Bayesian estimation which indicate that the estimator performs very well.

Table 5: Model A1: absolute bias(coverage)

Parameter	True Value	Bayes	MLO
β_1	.1	.00(.95)	.00(.97)
β_2	.3	.00(.94)	.01(.96)
β_3	.2	.01(.93)	.01(.94)
γ_{00}	.1	.00(.94)	.00(.94)
γ_{01}	.2	.01(.98)	.02(.98)
γ_{02}	.2	.00(.93)	.01(.91)

It is important to point out here why the ML and MLMC estimations have convergence problems. The interaction term $X_i Z_i$ in equation (42) is not an observed quantity. It is essentially $(X_{ij} - X_{.j})(Z_{ij} - Z_{.j})$ where X_{ij} and Z_{ij} are observed but $X_{.j}$ and $Z_{.j}$ are not observed. $X_{.j}$ and $Z_{.j}$ represent the true means of these variables in cluster j (or equivalently the random intercept effect) which are different from the sample means, i.e., the averages of the observations in the cluster. Because the likelihood for this model involves the product of two latent variables, it has no closed form expression and is computed through numerical integration. For the above model the Mplus implementation requires 5-dimensional integration which is very computationally demanding. To make the computation feasible the number of quadrature points per dimension is reduced to 4 with the ML estimation. That in turn leads to poor precision in the computation of the log-likelihood which eventually leads to non-convergence. Similarly the precision of the MLMC estimation is compromised as well.

The observed cluster averages $\overline{X_{.j}}$ and $\overline{Z_{.j}}$ are measurements for the true cluster means $X_{.j}$, and $Z_{.j}$, which have measurement error. The smaller the cluster size the bigger the measurement error. If that measurement error is not accounted for the regression coefficients can be biased. That bias is generally referred to as Lüdtke's bias, see Lüdtke et al. (2008) and Asparouhov and Muthén (2019). The bias occurs when there is a contextual effect in the model and the cluster sizes are relatively small, i.e., less than 50. If there is no contextual effect or the contextual effect is small or if the cluster sizes are large, the bias does not occur and in such situations it is safe to use the sample cluster average in place of the the true mean in the moderation model. If we replace $X_{.j}$ with the cluster average $\overline{X_{.j}}$ and $Z_{.j}$ with the cluster average $\overline{Z_{.j}}$, all the covariates in the model X_i, Z_i, X_iZ_i, $X_{.j}$, and $Z_{.j}$ become observed and the above model is essentially a simple univariate two-level regression which is very easy to estimate. Let's call this estimation method the MLO (maximum likelihood with observed centering).

In the above example, the contextual effect for X_{ij} and Z_{ij} is small because β_1 is close to γ_{01} and β_2 is close to γ_{02}. Therefore we can expect that the MLO method performs well for this example. The results for the MLO method are also included in Table 5 and we can see that indeed the method works well. It yields fast convergence in all cases and the parameter estimates and standard errors are satisfactory.

We illustrate Lüdtke's bias in the above model with one additional simulation study. We use parameters that are different from those specified in Preacher et al. (2016) montecarlo setups so that the variables have contextual effect. For this simulation study we generate 100 data sets with 500 clusters of size 10 using the following parameters $\beta_1 = .1$, $\beta_2 = -.6$, $\beta_3 = .6$, $\gamma_{00} = .1$, $\gamma_{01} = .7$, $\gamma_{02} = .9$, $Var(u_{0j}) = Var(\varepsilon_{ij}) = .7$. The means of $X_{.j}$ and $Z_{.j}$ are set to 0, the variances to .7 and the covariance to .1. The means of X_i and Z_i are set to 0, the variances to 2.7 and the covariance to 1.5. The results of the simulation study are presented

Table 6: Lüdtke's bias in model A1: absolute bias(coverage)

Parameter	True Value	Bayes	MLO	MLRO
β_1	.1	.00(.84)	.00(.82)	.00(.96)
β_2	-.6	.00(.90)	.00(.72)	.00(.99)
β_3	.6	.01(.93)	.00(.72)	.00(.88)
γ_{00}	.1	.00(.88)	.09(.59)	.09(.60)
γ_{01}	.7	.01(.95)	.29(.00)	.29(.00)
γ_{02}	.9	.01(.93)	.43(.00)	.43(.00)

in Table 6. For this simulation we also include the MLRO estimation, which is the same as MLO point estimation plus robust Huber-White sandwich standard errors.

The results show that the Bayes estimator performs well while both MLO and MLRO perform poorly due to Lüdtke's bias for all between level parameters, which also results in poor coverage. On the within level the results are unbiased for MLO but the standard errors are underestimated which results in poor coverage. In that respect MLRO is better as it resolves this issue but only for the within level parameters. Underestimation of the standard errors appears to be a problem unique to the moderation models. This problem does not occur with standard path analysis models, see Table 3 in Asparouhov and Muthén (2019). It is also important to note here that if random regression slopes are included in the model, the within level parameters β_i may also be biased, see Table 4 in Asparouhov and Muthén (2019). This within level bias will carry over to the moderation models as well.

Despite the fact that the MLO/MLRO estimators could be biased, we recommend that these estimators be used as a part of any moderation analysis.

MLO/MLRO can be used as a first preliminary step, which can be followed by the Bayesian estimation. The simplicity of the MLO/MLRO estimation is a very desirable attribute that no other estimation can match. This is why the approach should not be dismissed when it is available, i.e., when there is no missing data for the predictor and mediator. In addition, if the cluster sizes are 100 or more, these estimators can be used as the main method of estimation since they offer more options for model testing such as AIC/BIC and LRT.

3.3.2 Model A2: [Between part of L1 moderator] x [Within part of L1 predictor] (cross-level interaction)

The model is given by the following equations

$$X_{ij} = X_i + X_{.j} \tag{44}$$

$$Z_{ij} = Z_i + Z_{.j} \tag{45}$$

$$Y_{ij} = \beta_{0j} + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_{.j} + \varepsilon_{ij} \tag{46}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} X_{.j} + \gamma_{02} Z_{.j} + u_{0j} \tag{47}$$

We generate 100 data sets with 100 clusters of size 10. We analyze the data with the Bayes estimator, the ML estimator with numerical integration and 4 integration points per dimension and the ML estimator with Montecarlo (MLMC) integration with 500 integration points. The ML and MLMC did not converge. The results for the Bayes estimator are presented in Table 7. The Bayes estimator performs very well. It is important to note here that the Bayes estimator needs an additional option for this model: *variance=0.01;*. The role of this option is to

Table 7: Model A2: absolute bias(coverage)

Parameter	True Value	Bayes
β_1	.1	.00(.97)
β_2	.3	.00(.97)
β_3	.2	.01(.91)
γ_{00}	.1	.01(.94)
γ_{01}	.2	.00(.99)
γ_{02}	.2	.01(.96)

prevent slow/poor mixing due to residual variances fixed to 0.

3.3.3 Model A3: [Between part of L1 moderator] x [Between part of L1 predictor]

The model is given by the following equations

$$X_{ij} = X_i + X_{.j} \tag{48}$$

$$Z_{ij} = Z_i + Z_{.j} \tag{49}$$

$$Y_{ij} = \beta_{0j} + \beta_1 X_i + \beta_2 Z_i + \varepsilon_{ij} \tag{50}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} X_{.j} + \gamma_{02} Z_{.j} + \gamma_{03} X_{.j} Z_{.j} + u_{0j}. \tag{51}$$

We generate 100 data sets each containing 100 clusters of size 10. We analyze the data with the Bayes estimator, the ML estimator with numerical integration and 4 integration points per dimension and the ML estimator with Montecarlo (MLMC) integration with 1000 integration points. The results are presented in Table 8. All three estimators performed well in this situation. The Bayes estimator is 10 times

Table 8: Model A3: absolute bias(coverage)

Parameter	True Value	Bayes	ML	MLMC
β_1	.1	.00(.95)	.00(.97)	.00(.96)
β_2	.3	.00(.94)	.01(.93)	.00(.95)
γ_{00}	.1	.00(.94)	.01(.94)	.01(.89)
γ_{01}	.2	.01(.98)	.02(.93)	.02(.95)
γ_{02}	.2	.00(.93)	.00(.98)	.00(.98)
γ_{03}	.2	.00(.93)	.02(.91)	.00(.90)

faster than the ML estimator and 30 times faster than the MLMC estimator and takes less than a second for each replication.

3.3.4 Model A1 and A2 combination

The model is given by the following equations

$$X_{ij} = X_i + X_{.j} \tag{52}$$

$$Z_{ij} = Z_i + Z_{.j} \tag{53}$$

$$Y_{ij} = \beta_{0j} + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + \beta_4 X_i Z_{.j} + \varepsilon_{ij} \tag{54}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} X_{.j} + \gamma_{02} Z_{.j} + u_{0j}. \tag{55}$$

We generate 100 data sets with 100 clusters of size 10. We analyze the data with the Bayes, ML and MLMC estimators. The ML and MLMC estimators did not converge. The results for the Bayes estimator are presented in Table 9. The Bayes estimator performs well for this model as well.

Table 9: Model A1 plus A2 combination: absolute bias(coverage)

Parameter	True Value	Bayes
β_1	.1	.00(.97)
β_2	.3	.00(.94)
β_3	.2	.00(.94)
β_4	.2	.01(.94)
γ_{00}	.1	.01(.91)
γ_{01}	.2	.00(.95)
γ_{02}	.2	.02(.95)

3.3.5 Model B1: [L2 moderator] x [Within part of L1 predictor] (cross-level interaction)

In this model the moderator Z_{ij} is assumed to be a between level variable, i.e., $Z_{ij} = Z_j$. The model is given by the following equations

$$X_{ij} = X_i + X_{.j} \tag{56}$$

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_i + \beta_3 X_i Z_j + \varepsilon_{ij} \tag{57}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} X_{.j} + \gamma_{02} Z_j + u_{0j} \tag{58}$$

Note that unlike in the previous models, here we have a random slope β_{1j}. The model can be estimated with a non-random slope β_1 but we are following the fifth example from Preacher et al. (2016) where the model is with a random slope.

The presence of the random slope, however, complicates the identification of the model. In the earlier draft of Preacher et al. (2016) the model was specified as an unidentified model but in the current version of the supplemental materials the issue is resolved by a model modification. We illustrate these complications by considering first the simple situation where the random slope β_{1j} is regressed on Z_j. That is, we augment the above model with the equation

$$\beta_{1j} = \gamma_{10} + \gamma_{12} Z_j + u_{1j}. \tag{59}$$

If equation (59) is substituted in equation (57) we can clearly see that the coefficients γ_{12} and β_3 play the same role, i.e, these are the regression coefficient for the interaction term $X_i Z_j$. Thus both of these coefficients cannot be identified at the same time. Only one of the two coefficients can be present in the model.

An alternative way that this unidentification can appear in the model is as follows. Instead of estimating the regression equation (59) it is possible to estimate the variance covariance structure for the random effect β_{1j} and Z_j, which includes the covariance parameter. Such a model, however, is a reparameterization of (59). Thus we conclude that model B1 of Preacher et al. (2016) must have the covariance parameter between β_{1j} and Z_j fixed to 0. If the covariance parameter is not fixed to 0 the model would be unidentified because it would includes the two essentially equivalent parameters: the covariance between β_{1j} and Z_j as well as β_3. One of these two parameters must be fixed to zero.

In the following simulation we estimate the above model, assuming that the covariance between β_{1j} and Z_j as well as the covariance between β_{1j} and $X_{.j}$ are not estimated, i.e., these covariance parameters are fixed to zero. Equivalently, we can assume that γ_{12} in equation (59) is fixed to 0. The covariance between Z_j and $X_{.j}$ is estimated. We generate 100 data sets with 100 clusters of size 10. The ML and the MLMC estimation methods did not converge for this model and thus we report the results in Table 10 only for the Bayes estimator. The Bayes

Table 10: Model B1: absolute bias(coverage)

Parameter	True Value	Bayes
β_3	.2	.00(.92)
γ_{00}	.1	.01(.92)
γ_{01}	.2	.01(.94)
γ_{02}	.2	.00(.98)
γ_{10}	.1	.01(.96)

estimator performs well also for this example.

3.3.6 Model B2: [L2 moderator] x [Between part of L1 predictor]

In this model the moderator is again assumed to be a between level variable Z_j. The model is given by the following equations

$$X_{ij} = X_i + X_{.j} \tag{60}$$

$$Y_{ij} = \beta_{0j} + \beta_1 X_i + \varepsilon_{ij} \tag{61}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} X_{.j} + \gamma_{02} Z_j + i + \gamma_{03} X_{.j} Z_j + u_{0j}. \tag{62}$$

We generate 100 data sets with 100 clusters of size 10. The data is analyzed by the three estimators Bayes, ML, MLMC and the results are reported in Table 11. All three estimators performed well for this model.

Table 11: Model B2: absolute bias(coverage)

Parameter	True Value	Bayes	ML	MLMC
β_1	.2	.01(.93)	.00(.94)	.00(.95)
γ_{00}	.1	.01(.99)	.01(.95)	.01(.93)
γ_{01}	.2	.01(.95)	.01(.93)	.01(.91)
γ_{02}	.2	.02(.92)	.00(.93)	.00(.89)
γ_{03}	.2	.01(.99)	.01(.97)	.02(.87)

3.3.7 Model A1 with random slope for the interaction term

The model can be described by the following equations

$$X_{ij} = X_i + X_{.j} \tag{63}$$

$$Z_{ij} = Z_i + Z_{.j} \tag{64}$$

$$Y_{ij} = \beta_{0j} + \beta_1 X_i + \beta_2 Z_i + \beta_{3j} X_i Z_i + \varepsilon_{ij} \tag{65}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} X_{.j} + \gamma_{02} Z_{.j} + u_{0j} \tag{66}$$

$$\beta_{3j} = \gamma_{30} + u_{3j}. \tag{67}$$

We conduct a simulation study again using 100 samples with 100 clusters of size 10. Only the Bayes estimator converged in this case. The results are presented in Table 12. The Bayes estimator performs well for this model as well.

Table 12: Model A1 with random slope: absolute bias(coverage)

Parameter	True Value	Bayes
β_1	.1	.00(.94)
β_2	.3	.00(.97)
γ_{00}	.1	.01(.96)
γ_{01}	.2	.02(.96)
γ_{02}	.2	.00(.97)
γ_{30}	.2	.02(.96)

3.3.8 Model C: [L2 moderator] x [L2 predictor]

In this model the moderator and the predictor are between level variable Z_j and X_j respectively. The model is given by the following equations

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij} \tag{68}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} X_j + \gamma_{02} Z_j + i + \gamma_{03} X_j Z_j + u_{0j}. \tag{69}$$

If the variables X_j and Z_j have no missing values then the model can estimated as a regular two-level model (without the Mplus moderation command XWITH). The interaction term $X_j Z_j$ can be computed with the Mplus DEFINE command and be treated just like any other covariate. If there are missing data for these variables, however, the Mplus DEFINE command can not be used to simply multiply the two variables, and instead the moderation model estimation has to be utilized.

In this simulation we generate 100 data sets with 100 clusters of size 10. Missing data for the variable X_j is generated as follows

$$Prob(X_j \text{is missing}) = \frac{1}{1 + Exp(Z_j)}. \tag{70}$$

Table 13: Model C: absolute bias(coverage)

Parameter	True Value	Bayes	ML	MLMC
γ_{00}	.1	.01(.90)	.01(.92)	.00(.94)
γ_{01}	.1	.00(.87)	.00(.93)	.00(.93)
γ_{02}	.2	.00(.98)	.00(.96)	.01(.95)
γ_{03}	.1	.01(.94)	.00(.94)	.01(.96)

This method of generating missing data is MAR (and not MCAR). Likelihood based estimators such as Bayes, ML and MLMC are guaranteed to produced unbiased estimates for such missing data mechanisms.

The results of this simulation are reported in Table 13. All three estimators performed well for this model.

3.4 Multilevel Moderated Mediation

The multilevel moderated mediation model is discussed in Zyphur et al. (2019) and is illustrated with examples. The estimation described in that article is based on the plausible value methodology so that the modeling uses estimates of the latent variable values. In this section we illustrate how the model can be estimated with Mplus V8.3 directly using the Bayes estimator. Suppose that Y_{ij}, M_{ij}, X_{ij} and Z_{ij} are the observed variables for individual i in cluster j. As usual the variables are decomposed as within and between

$$Y_{ij} = Y_i + Y_{.j} \tag{71}$$

$$M_{ij} = M_i + M_{.j} \tag{72}$$

$$X_{ij} = X_i + X_{.j} \tag{73}$$

$$Z_{ij} = Z_i + Z_{.j} \tag{74}$$

On the within and the between level the model is described as follows

$$M_i = \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + \varepsilon_{1,ij} \tag{75}$$

$$Y_i = \beta_4 X_i + \beta_5 Z_i + \beta_6 M_i + \beta_7 X_i Z_i + \beta_8 M_i Z_i + \varepsilon_{2,ij} \tag{76}$$

$$M_{.j} = \alpha_1 + \gamma_1 X_{.j} + \gamma_2 Z_{.j} + \gamma_3 X_{.j} Z_{.j} + \varepsilon_{3,j} \tag{77}$$

$$Y_{.j} = \alpha_2 + \gamma_4 X_{.j} + \gamma_5 Z_{.j} + \gamma_6 M_{.j} + \gamma_7 X_{.j} Z_{.j} + \gamma_8 M_{.j} Z_{.j} + \varepsilon_{4,j} \tag{78}$$

To illustrate the Bayesian estimation for the above model we conduct the following simulation study. We generate 100 data sets with 100 clusters of size 10. The data is generated using the above model and the following parameters values. The values of α_i, β_i and γ_i are given in Table 14. The variance of the residual variables $\varepsilon_{1,ij}$, $\varepsilon_{2,ij}$, $\varepsilon_{3,j}$, $\varepsilon_{4,j}$ are set to 0.7. The variance of X_i, Z_i, $X_{.j}$, $Z_{.j}$ are also set to 0.7. The covariance between X_i and Z_i is set to .1. The covariance between $X_{.j}$ and $Z_{.j}$ is set to .1. The means of $X_{.j}$ and $Z_{.j}$ are set to 0. Table 14 contains the results when the above model is estimated with the Bayes estimator. The results indicate that the Bayes estimator performs very well.

3.5 Three way interactions

The Bayes model estimation described so far is specific for two-way interactions, i.e., the product of two variables. It is possible, however, using the same methodology to form 3-way and higher order interactions. Suppose that we want

Table 14: Multilevel Moderated Mediation Model: absolute bias(coverage)

Parameter	True Value	Bayes
α_1	.2	.00(.92)
α_2	.6	.00(.97)
β_1	.1	.01(.93)
β_2	.3	.00(.94)
β_3	.2	.00(.91)
β_4	.3	.00(.94)
β_5	.2	.00(.99)
β_6	.4	.00(.93)
β_7	.3	.01(.97)
β_8	.1	.00(.94)
γ_1	.2	.00(.97)
γ_2	.1	.00(.94)
γ_3	.3	.00(.95)
γ_4	.4	.00(.94)
γ_5	.1	.01(.96)
γ_6	.1	.01(.86)
γ_7	.2	.01(.92)
γ_8	.2	.01(.92)

to include a term $\eta_1\eta_2\eta_3$ as a predictor of a variable Z. We can accomplish that by using a new latent variables η_{12} and two two-way interactions as follows

$$\eta_{12} = \eta_1\eta_2 + \varepsilon_{12} \tag{79}$$

$$Z = \beta_2\eta_{12}\eta_3 + \varepsilon \tag{80}$$

If we fix the variance of ε_{12} to zero then $\eta_{12}\eta_3 = \eta_1\eta_2\eta_3$ and we have the desired 3 way interaction. With the Bayesian estimation, however, fixing the variance to 0 is not an option and thus we have to fix it to a small positive value, which makes the above model an approximation of the 3 way interaction model. The smaller the value is, the more precise the approximation but also the slower the mixing. In our experience, choosing a value that represents around 1% of the variance of Z works well. Alternatively, if the variances of η_1, η_2 and η_3 are set to 1 fixing the variance of ε_{12} to 0.01 would work well too.

We illustrate the three way interaction with the following single-level simulation study. Consider the following model with 9 observed variables and 3 latent variables

$$Y_i = \nu_i + \lambda_i\eta_1 + \varepsilon_i, i = 1, 2, 3 \tag{81}$$

$$Y_i = \nu_i + \lambda_i\eta_2 + \varepsilon_i, i = 4, 5, 6 \tag{82}$$

$$Y_i = \nu_i + \lambda_i\eta_3 + \varepsilon_i, i = 7, 8, 9. \tag{83}$$

The structural part of the model is given by the following equation which includes

Table 15: Three way interactions: absolute bias(coverage)

Parameter	True Value	Bayes
β_1	.5	.00(.94)
β_2	.7	.00(.92)
β_3	.4	.02(.89)

the three way interaction (cubic term) $\eta_1^2\eta_2$

$$\eta_3 = \beta_1\eta_1 + \beta_2\eta_2 + \beta_3\eta_1^2\eta_2 + \xi. \tag{84}$$

We generate 100 data sets of size 1000 using the following model parameters: $\alpha_i = 0$, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_7 = 1$, $\lambda_5 = \lambda_6 = .8$, $\lambda_8 = \lambda_9 = .9$, $\beta_1 = .5$, $\beta_2 = .7$, $\beta_3 = .4$, $Var(\eta_1) = Var(\eta_2) = Var(\varepsilon_i) = 1$, and $Var(\eta_3) = .5$. We analyze the data with the above model and the Bayesian estimator and present the results for a subset of the parameters in Table 15. The results indicate that the Bayesian estimation performs well.

Appendix: Standardization in matrix terms

This section describes standardization of the general model in Mplus when latent variable interactions are present. Suppose that Y is a vector of dependent variables, X is a vector of covariates, η is a vector of latent variables. All residuals are assumed normal.

Suppose that the variables (Y,η,X) are split into two disjoint set of variables V_1 and V_2. V_1 set of dependent variables that are not part of interaction terms and V_2 is a set of variables that are a part of interaction terms. Suppose that V_1 is a vector of size p_1 and V_2 is a vector of size p_2. The SEM equation is given by these two equations

$$V_1 = \alpha_1 + B_1 V_1 + C_1 V_2 + \sum_{i=1}^{k} D_i (V_{2,f(i)} V_{2,g(i)}) + \varepsilon_1$$

$$V_2 = \alpha_2 + B_2 V_2 + \varepsilon_2$$

where $\alpha_1, B_1, C_1, D_i, \alpha_2, B_2$ are model parameters. The vectors α_1, D_i are of length p_1 while the vector α_2 is of length p_2. The matrices B_1, C_1 and B_2 are of size $p_1 \times p_1$, $p_1 \times p_2$ and $p_2 \times p_2$ respectively.

The residual variable ε_1 has zero mean and variance covariance Θ and ε_2 has zero mean and variance covariance Ψ. The residuals ε_1 are not considered independent of the residuals ε_2. Let's call the covariance $F = Cov(\varepsilon_1, \varepsilon_2)$. The functions $f(i)$ and $g(i)$ simply define the interaction terms, i.e., $f(i)$ and $g(i)$ are integers between 1 and p_2 and k is the number of interaction terms in the model.

We can assume that all covariates X are in the V_2 vector and the V_1 vector consists only of η and Y variables that are regressed on interaction terms, while the remaining η and Y variables are in vector V_2. We can compute the model implied mean and variance for these variables as follows. For the variables V_2 we get

$$E(V_2) = \mu_2 = (1 - B_2)^{-1}\alpha_2$$

$$Var(V_2) = \Sigma_2 = (1 - B_2)^{-1}\Psi((1 - B_2)^{-1})^T$$

For V_1 we get

$$E(V_1) = (1-B_1)^{-1}\alpha_1 + (1-B_1)^{-1}C_1\mu_2 + (1-B_1)^{-1}\sum_{i=1}^{k} D_i(\mu_{2,f(i)}\mu_{2,g(i)} + \Sigma_{2,f(i),g(i)}).$$

Denote by

$$V_{20} = V_2 - \mu_2$$

$$\mu_{10} = (1 - B_1)^{-1}\alpha_1 + (1 - B_1)^{-1}C_1\mu_2 + (1 - B_1)^{-1}\sum_{i=1}^{k} D_i(\mu_{2,f(i)}\mu_{2,g(i)})$$

$$V_{10} = (1 - B_1)^{-1}C_1V_{20} + (1 - B_1)^{-1}\varepsilon_1 + (1 - B_1)^{-1}\sum_{i=1}^{k} D_i(\mu_{2,f(i)}V_{20,g(i)}) +$$

$$(1 - B_1)^{-1}\sum_{i=1}^{k} D_i(V_{20,f(i)}\mu_{2,g(i)}).$$

Then

$$V_1 = \mu_{10} + V_{10} + (1 - B_1)^{-1}\sum_{i=1}^{k} D_i(V_{20,f(i)}V_{20,g(i)}).$$

Another representation for V_{10} is

$$V_{10} = QV_{20} + (1 - B_1)^{-1}\varepsilon_1$$

where the matrix Q combines all the coefficients from the terms involving V_{20}.

The above equation is essentially the definition of Q. Note now that

$$Cov(\varepsilon_1, V_{20}) = F((1 - B_2)^{-1})^T$$

and thus

$$Var(V_{10}) = Q\Sigma_2 Q^T + (1 - B_1)^{-1}\Theta((1 - B_1)^{-1})^T + Q(1 - B_2)^{-1}F^T((1 - B_1)^{-1})^T +$$

$$(1 - B_1)^{-1}F((1 - B_2)^{-1})^T Q^T.$$

Using the fact that the covariance between $V_{20,f(i)}V_{20,g(i)}$ and V_{20} and the covariance between $V_{20,f(i)}V_{20,g(i)}$ and ε_1 are zero we get that

$$Var(V_1) = Var(V_{10}) + \sum_{i,j} D_i Cov(V_{20,f(i)}V_{20,g(i)}, V_{20,f(j)}V_{20,g(j)})D_j^T =$$

$$Var(V_{10}) + \sum_{i,j} D_i D_j^T(\Sigma_{2,f(i),f(j)}\Sigma_{2,g(i),g(j)} + \Sigma_{2,f(i),g(j)}\Sigma_{2,g(i),f(j)}).$$

Note also that

$$Cov(V_1, V_2) = Cov(V_{10}, V_{20}) = Q\Sigma_2 + (1 - B_1)^{-1}F((1 - B_2)^{-1})^T.$$

References

[1] Anderson, T. W. (1984). An Introduction to multivariate statistical analysis (2nd ed.). New York: Wiley.

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

[3] Asparouhov, T & Muthén, B. (2010b). Plausible values for latent variable using Mplus. http://statmodel.com/download/Plausible.pdf

[4] Asparouhov, T. & Muthén, B. (2019). Latent variable centering of predictors and mediators in multilevel and time-series models. Structural Equation Modeling: A Multidisciplinary Journal, 26(1), 119-142.

[5] Arminger, G., & Muthén, B. (1998). A Bayesian approach to nonlinear variable models using the Gibbs sampler and the Metropolis–Hastings algorithm. Psychometrika, 63, 271-310.

[6] Cudeck, R., Harring, J. R., & du Toit, S. H. C. (2009). Marginal maximum likelihood estimation of a latent variable model with interaction. Journal of Educational and Behavioral Statistics, 34, 131-144.

[7] Klein, A. & Moosbrugger, H. (2000) Maximum likelihood estimation of latent interaction effects with the LMS method. Psychometrika, 65, 457-474.

[8] Klein, A. G., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. Multivariate Behavioral Research, 42, 647-673.

[9] Lüdtke, O., Marsh, H.W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. Psychological Methods, 13, 203-229.

[10] Marsh, H. W., Wen, Z., & Hau, K. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. Psychological Methods, 9, 275-300.

[11] Marsh, H. W., Nagengast, B., & Morin, A. J. (2013). Measurement invariance of Big-Five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. Developmental Psychology, 49, 1194-1218.

[12] Mooijaart, A. & Bentler, P. (2010). An alternative approach for nonlinear latent variable models. Structural Equation Modeling, 17, 357-373.

[13] Mooijaart, A. & Satorra, A. (2009). On Insensitivity of the Chi-Square Model Test to Nonlinear Misspecification in Structural Equation Models. Psychometrika, 74, 443-455.

[14] Muthén, B. & Asparouhov, T. (2003) Modeling interactions between latent and observed continuous variables using maximum-likelihood estimation in Mplus. Mplus Web Notes, #6. http://www.statmodel.com/download/webnotes/webnote6.pdf.

[15] Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. Psychological Methods, 17(3), 313-335.

[16] Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016) Multilevel structural equation models for assessing moderation within and across levels of analysis. Psychological Methods, 21, 189-205. Suplemental materials at http://quantpsy.org/pubs/preacher_zhang_zyphur_2016_(code.appendix).pdf

[17] Satorra, A. (1990). Robustness issues in structural equation modeling: a review of recent developments. Quality & Quantity, 24, 367-386.

[18] Satorra, A. (2002). Asymptotic robustness in multiple group linear-latent variable models. Econometric Theory, 18, 297-312.

[19] Wen, Z., Marsh, H. W., & Hau, K.-T. (2010). Structural equation models of latent interactions: An appropriate standardized solution and its scale-free properties. Structural Equation Modeling, 17, 1-22.

[20] Zyphur, M. J., Zhang, Z., Preacher, K. J., & Bird, L. J. (2019). Moderated mediation in multilevel structural equation models: Decomposing effects of race on math achievement within versus between high schools in the United States. In S. E. Humphrey & J. M. LeBreton (Eds.), The handbook of multilevel theory, measurement, and analysis (pp. 473-494). Washington, DC, US: American Psychological Association.