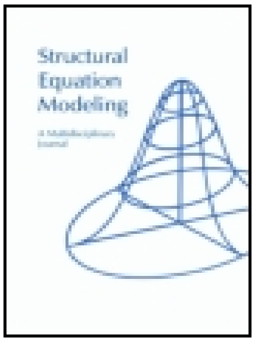

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: http://www.tandfonline.com/loi/hsem20

Latent Variable Centering of Predictors and Mediators in Multilevel and Time-Series Models

Tihomir Asparouhov & Bengt Muthén

Structural Equation Modeling: A Multidisciplinary Journal, 00: 1–24, 2018

ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2018.1511375

Routledge
Taylor & Francis Group

Check for updates

TEACHER'S CORNER

Latent Variable Centering of Predictors and Mediators in Multilevel and Time-Series Models

Tihomir Asparouhov and Bengt Muthén
Mplus, Los Angeles, CA

INTRODUCTION

In hierarchical linear regression models the question regarding the proper way to construct covariates has long been the focus of attention. The main issue is whether or not the covariates should be group mean centered, grand mean centered, uncentered or latent mean centered. Group mean centering has been recommended in Raudenbush and Bryk (2002) and Enders and Tofighi (2007) for example. Grand mean centering has al so been recommended for particular situations in Enders and Tofighi (2007), while the uncentered method has been considered in Hamaker and Grasman (2015) in special applications in the time-series context. Latent mean centering has been used in Lüdtke et al. (2008), Asparouhov and Muthén (2006a) and Preacher, Zyphur, and Zhang (2010). The latent mean centering has also been the main modeling approach in multilevel structural equation models, see Muthén (1994). In the time-series context, the latent mean centering has been used in Asparouhov, Hamaker, and Muthén (2018) as a way to resolve Nickell's bias, see Nickell (1981). The latent mean centering has also been used in Preacher, Zhang, and Zyphur (2016) in the context of multilevel moderation models. In the context of three level modeling, the latent mean centering has been utilized in Marsh, Kuyper, Morin, Parker, and Seaton (2014). A hybrid centering method, which is based on a combination of the latent mean centering and the uncentered method, has been used in multilevel mediation models with random slopes, see Preacher et al. (2010).

Despite the long and well established history of multilevel modeling, the covariate centering issue remains unsettled. There are three aspects that generally drive the consideration of centering: interpretability, quality of the statistical methodology in terms of bias and MSE (mean squared error) of the estimation, as well as software availability. In addition to that there are three other aspects that play an important role in the discussion: the presence of missing data in the covariate, whether or not the slope in front of the covariate is random and whether or not the dependent and the predictor variables are categorical or continuous. The purpose of this note is to further explain the case for the latent mean centering as the most accurate, most easily interpretable and most widely applicable method. Despite the fact that the latent centering method has been successfully utilized in the past to resolve a variety of estimation problems, the method has not been easily accessible in the most general multilevel model, for example, for models with random slopes. Enders and Tofighi (2007) discuss centering for multilevel models with random slopes, however, the latent centering method is not considered at all. In Preacher et al. (2010), mediation models with random slopes use the uncentered estimation method which is subsequently reparameterized so that implications can be made for the centered model.

In this article, we focus on the centering options for a predictor with a random slope. In Mplus Version 8.1, the latent centering method has been extended for the general multilevel model with random slopes using Bayesian estimation. Because of that the latent centering can now be used in many situations where previously no estimation method was easily available. We can now also conduct simulation studies that reveal the disadvantages of the more traditional methods of observed group mean centering, grand mean centering, the uncentered and the hybrid methods. Several such simulation studies are described below.

The maximum-likelihood estimation of the latent centering model with a random slope is in principle possible,

Correspondence should be addressed to Tihomir Asparouhov.
E-mail: tihomir@statmodel.com
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/HSEM.

using numerical integration, but is impractical and limited. The log-likelihood for this model does not allow for a closed form expression because it includes the product/interactions of between level random effects. Within the Mplus ML framework all the random effects will need to be numerically integrated and thus such an estimation is limited by the number of variables, covariates and random effects. With more than three or four random effects the ML estimation based on numerical integration will be slow, less precise and quite likely to lead to convergence problems. On the other hand, the observed group mean centering, the grand mean centering and the uncentered method all have closed form likelihood expressions, are generally easy to estimate and are not limited to the number of variables, covariates and random effects. Because of that the Mplus default for estimating multi-level models with random slopes has been to use the uncentered approach, but the group mean and grand mean centering are easily accessible as well. With the new development in the Bayesian estimation algorithms in Mplus Version 8.1, the playing field for the different centering methods is finally leveled and we can now easily estimate and compare these centering options.

If a covariate in the multilevel model with random slope has missing values there are three implications. First, the ML estimation method within Mplus will again require numerical integration due to the multiplication of latent variables. That in turn will limit the scope of the model in terms of the dimensions of numerical integration and to a large extent the ML estimation becomes impractical even when we use the three traditional centering methods: group mean centering, grand mean centering or the uncentered method. In Mplus Version 8 it is possible, however, to estimate the model with the observed group mean centering method using Bayesian estimation even in the presence of missing data. The second implication concerns specifically the observed group mean centering. This centering amounts to subtracting the average covariate value in a particular cluster from the covariate values in that cluster. The problem that arises in the presence of missing data is that unless the missing data is missing completely at random (MCAR) the sample mean is not the actual mean for the covariate. Therefore, we should not expect good estimation performance when using observed group mean centering in the presence of missing data, even if a fast and accurate estimation method exists. The fault in the approach occurs before the estimation begins, i.e., in the centering of the data with the wrong means. The third aspect of having missing data in the covariate is that the group mean centering is no longer equivalent to the latent mean centering for applications with large cluster. It is well known that if the cluster sizes in the data are greater than 100, the group mean centering method is practically equivalent to the latent mean centering. That is because the error in the sample mean estimate of the true cluster mean decreases as the cluster sample size increases. Cluster sizes of 100 or more are sufficient in most cases to expect little difference between the methods. This argument, however, breaks down in the presence of missing data, again, because the cluster sample mean is not necessarily consistent estimate of the true mean if the missing data is not MCAR.

THE DIFFERENT CENTERING OPTIONS

We illustrate the different centering options using a simple two-level regression model. Let Y_{ij} be a dependent variable for individual i in cluster j and let X_{ij} be the corresponding covariate. The standard two-level regression formulation, see Raudenbush and Bryk (2002), is usually given as follows:

$$Y_{ij} = \alpha_j + \beta_{1j}(X_{ij} - \overline{X}_{.j}) + \varepsilon_{w,ij} \tag{1}$$

$$\alpha_j = \alpha + \beta_2 \overline{X}_{.j} + \varepsilon_{b,j} \tag{2}$$

$$\beta_{1j} = \beta_1 + \xi_j. \tag{3}$$

The model has two random effects: the random intercept α_j and the random slope β_{1j}. Typically, the covariate in equation (1) is centered by its cluster specific mean $\overline{X}_{.j}$ which is then used in (2) as a cluster level predictor for the random intercept.

The above model allows the effect of the covariate on the within level to be different from the effect of the covariate on the between level. This is important as these effects are often not equal. If we estimate just one effect using the uncentered model,

$$Y_{ij} = \alpha_j + \beta_{0j} X_{ij} + \varepsilon_{w,ij} \tag{4}$$

$$\alpha_j = \alpha + \varepsilon_{b,j} \tag{5}$$

$$\beta_{0j} = \beta_0 + \xi_j. \tag{6}$$

the average effect β_0 becomes an "uninterpretable blend" of the within and the between effects β_1 and β_2. The blend is approximately

$$\beta_0 \approx \frac{w_1\beta_1 + w_2\beta_2}{w_1 + w_2}, \tag{7}$$

where the weights are $w_1 = 1/Var(\hat{\beta}_1)$, $w_2 = 1/Var(\hat{\beta}_2)$, see Raudenbush and Bryk (2002).

The model in equations (1–3) can also be described in terms of the following two-level decomposition. The two variables Y_{ij} and X_{ij} can be decomposed as within and between components as follows:

$$Y_{ij} = Y_{w,ij} + Y_{b,j} \tag{8}$$

$$X_{ij} = X_{w,ij} + X_{b,j} \tag{9}$$

Here $Y_{b,j}$ and $X_{b,j}$ are the cluster specific contributions to these variables while $Y_{w,ij}$ and $X_{w,ij}$ are the individual specific contributions. We can also interpret $Y_{w,ij}$ and $X_{w,ij}$ as the zero-mean residuals in the random intercept two-level regressions. The variables $Y_{b,j}$ and $X_{b,j}$ can be interpreted as the cluster specific means, i.e.,

$$Y_{b,j} = E(Y_{ij}|j) \tag{10}$$

$$X_{b,j} = E(X_{ij}|j). \tag{11}$$

The relationship between the variables Y and X can be expressed as a two-level linear regression model as follows:

$$Y_{w,ij} = \beta_{1j}X_{w,ij} + \varepsilon_{w,ij} \tag{12}$$

$$Y_{b,j} = \alpha_j = \alpha + \beta_2 X_{b,j} + \varepsilon_{b,j} \tag{13}$$

$$\beta_{1j} = \beta_1 + \xi_j. \tag{14}$$

In the above model α, β_1 and β_2 are non-random parameters. The random effect β_{1j} is the random slope on the within level and can be identified as a cluster specific effect because we have multiple individuals in cluster j. This random slope could also be estimated as a non-random fixed coefficient which is invariant across all the cluster but such an assumption is generally not needed and in most situation is overly restrictive. Note also that the within level model does not have a regression intercept. Such an intercept would not be identified and because the within level components are assumed to have zero mean that coefficient is assumed zero. In the above model we treat the covariate X as a predictor variable both on the within and the between level. The residual variable $\varepsilon_{w,ij}$ is assumed normally distributed with $N(0, \sigma_w)$ distribution, while the between level residual $\varepsilon_{b,j}$ and ξ_j are assumed to have bivariate normal zero mean distribution with variance covariance matrix Σ.

The above model has a clear separation of the effects as cluster level effects and as within cluster effects. It is the most basic two-level linear regression model that can be constructed using two variables. The one fundamental question that arises immediately is how to obtain or how to treat $Y_{b,j}$ and $X_{b,j}$. In standard multilevel regression models such as those described in Raudenbush and Bryk (2002) the two variables are generally treated differently. The variable $Y_{b,j}$ is treated as unknown and as a random effect, while $X_{b,j}$ is usually treated differently. While the treatment of $Y_{b,j}$ as a random effect is largely agreed upon, the treatment of $X_{b,j}$ remains unsettled and is referred to as the centering issue. In the following sections we describe the five different treatments of $X_{b,j}$: the observed group mean centering, the latent mean centering, the observed grand mean centering, the uncentered method and the hybrid method.

The observed group mean centering

This approach replaces the cluster mean $X_{b,j}$ with the cluster sample mean $\overline{X}_{.j}$, i.e., with the average of all observations in cluster j. The model simplifies substantially as $X_{w,ij}$ and $X_{b,j}$ are treated as observed and known quantities. This approach, however, suffers from the fact that the sample mean $\overline{X}_{.j}$ is not the true mean $X_{b,j}$, and the error in that estimate is not accounted for.

More explicitly, the observed group mean centering approach estimates the following model:

$$Y_{ij} = Y_{w,ij} + Y_{b,j} \tag{15}$$

$$Y_{w,ij} = \beta_{1j}(X_{ij} - \overline{X}_{.j}) + \varepsilon_{w,ij} \tag{16}$$

$$Y_{b,j} = \alpha + \beta_2 \overline{X}_{.j} + \varepsilon_{b,j} \tag{17}$$

$$\beta_{1j} = \beta_1 + \xi_j. \tag{18}$$

or if we combine the first three equations the model can be expressed as follows:

$$Y_{ij} = \alpha + \beta_2 \overline{X}_{.j} + \beta_{1j}(X_{ij} - \overline{X}_{.j}) + \varepsilon_{b,j} + \varepsilon_{w,ij}. \tag{19}$$

The model is identical to the model given in equations (1–3).

The latent mean centering

The latent mean centering approach treats $X_{b,j}$ as an unknown quantity that has to be estimated and thus properly accounts for the measurement error. The model can be expressed as follows:

$$Y_{ij} = Y_{w,ij} + Y_{b,j} \tag{20}$$

$$X_{ij} = X_{w,ij} + X_{b,j} \tag{21}$$

$$Y_{w,ij} = \beta_{1j}X_{w,ij} + \varepsilon_{w,ij} \tag{22}$$

$$Y_{b,j} = \alpha + \beta_2 X_{b,j} + \varepsilon_{b,j} \tag{23}$$

$$\varepsilon_{w,ij} \sim N(0, \sigma_1), X_{w,ij} \sim N(0, \sigma_{w,x}) \tag{24}$$

$$\beta_{1j} \sim N(\beta_1, \sigma_2), \varepsilon_{b,j} \sim N(0, \sigma_3), X_{b,j} \sim N(\mu_x, \sigma_{b,x}). \tag{25}$$

Note that in this model we estimate three random effects: β_{1j}, $Y_{b,j}$ and $X_{b,j}$, while in the observed group mean centering approach there are only two random effects β_{1j} and $Y_{b,j}$. The latent mean centering is also sometimes referred to as the latent covariate approach because the covariates in equations (22–23) are both latent variables. If the cluster size is sufficiently large, i.e., greater than 100, and there is no missing data in the covariate X_{ij} the error in the sample mean estimate $\overline{X}_{.j}$ would be negligible in most applications and thus we can expect that the latent mean centering and the observed group mean centering methods would yield similar results.

It is worth noting here that in certain applications the observed group mean centering might be more appropriate than the latent mean centering. For example, if all the units within the cluster, for example, geographical cluster, are sampled then the sample mean really represents the actual mean, i.e., it does not have an error. In that case, the latent mean centering approach would be considered inferior because it assumes that the cluster means are unknown quantities measured with error. The latent centering approach generally assumes that the cluster sizes are large (or infinite) and a small part of the clusters are actually sampled. In principle it is safe to assume that if less than 5% of the cluster population is sampled, the latent variable assumption is satisfied. If, however, the cluster sizes are finite and the proportion of sampled units are between 5% and 100% neither the latent variable centering nor the observed group centering assumptions are appropriate and both would be approximations. Complex survey methodology dealing with finite population sampling may provide the most realistic approach in such situations, see Asparouhov and Muthén (2006b). In many common applications, however, particularly in the social sciences, the clusters units themselves are sampled from a large population. Even if data is subsequently collected from all the members of the cluster, the cluster sample mean should not be treated as the actual mean. The actual mean is the mean of the large population where the cluster units were sampled from. The cluster sample mean represents the mean for the members selected in the cluster and it does not represent the mean of the units in the large population that could have been selected in the cluster. In such situations, the latent mean centering would clearly be more appropriate than the observed mean centering as it accounts for the sampling error in the mean estimate.

Uncentered method

The uncentered approach is the third centering option. With this method we estimate the model,

$$Y_{ij} = \alpha + \beta_2 \overline{X}_{.j} + \beta_{1j} X_{ij} + \varepsilon_{b,j} + \varepsilon_{w,ij} \tag{26}$$

$$\beta_{1j} = \beta_1 + \xi_j \tag{27}$$

i.e., the random slope multiplies the full covariate value instead of the within only part. Note that this model can be re-written as:

$$Y_{ij} = \alpha + (\beta_2 + \beta_{1j})\overline{X}_{.j} + \beta_{1j}(X_{ij} - \overline{X}_{.j}) + \varepsilon_{b,j} + \varepsilon_{w,ij} \tag{28}$$

i.e., this model is similar to the group mean centering approach with the exception that the coefficient in front of $\overline{X}_{.j}$ is now $\beta_2 + \beta_{1j}$. The model can also be written as:

$$Y_{ij} = Y_{w,ij} + Y_{b,j} \tag{29}$$

$$Y_{w,ij} = \beta_{1j}(X_{ij} - \overline{X}_{.j}) + \varepsilon_{w,ij} \tag{30}$$

$$Y_{b,j} = \alpha + (\beta_2 + \beta_{1j})\overline{X}_{.j} + \varepsilon_{b,j} \tag{31}$$

$$\beta_{1j} = \beta_1 + \xi_j. \tag{32}$$

If $Var(\xi_j) = 0$, then the random slope is a fixed slope and the uncentered model is simply a reparameterization of the group mean centered model. If both models are estimated with the ML estimator the log-likelihood values will be identical. If, however, $Var(\xi_j)>0$ the uncentered model is not a reparameterization of the group mean centering model. In fact, we can see that the uncentered model would be a poor approximation for the observed or latent centering models. This is because in equation (31) there is a random effect term β_{1j} that should not be there. This random effect is predominately determined by the within level model, and is designed to help the within level model find the best fitting linear regression for each cluster. Confounding the random effect with the between level regression slope as this model shows would prevent the between level part of the model to find the most optimal linear regression equation. In equation (31), we see that in an artificial way the interaction term $\beta_{1j}\overline{X}_{.j}$ becomes a predictor of $Y_{b,j}$. Not only is this interaction term added but it is added with a constant regression coefficient of 1. The standard method for adding an interaction terms in regression analysis would be to include a regression coefficient that is to be estimated, for example, the interaction term could be added as a $\beta_3\beta_{1j}\overline{X}_{.j}$ where the β_3 coefficient is free to be estimated and if the interaction term is not needed the coefficient will be estimated to zero. Instead the above model fixes that interaction effect coefficient to 1, which is an unreasonable assumption. Note also that in general the residual variables $\varepsilon_{b,j}$ and ξ_j can be correlated and such a correlation can be estimated. Alternatively, instead of estimating such a correlation the variable β_{1j} can be used as a predictor for $Y_{b,j}$. Thus, the above model would include as predictors for $Y_{b,j}$ the variables $\overline{X}_{.j}$, β_{1j} and the interaction term $\beta_{1j}\overline{X}_{.j}$ and the coefficients in front of $\overline{X}_{.j}$ and β_{1j} would be estimated based on the information in the data while the coefficient in front of the interaction term will be forced to be 1 without allowing the data to determine that coefficient. Preacher et al. (2016), see footnote 2, also argue that the interaction term $\beta_{1j}\overline{X}_{.j}$ "may be difficult to justify". Thus in general we can expect to see problems when the uncentered model is used as an approximation for the latent or the observed group mean centering models.

The grand mean centering

The grand mean centering is the fourth centering alternative for the above model. It simply amounts to subtracting the total sample average from the covariate instead of the cluster sample average. Since this amounts to nothing more but a scale shift in the covariate, the model will essentially be equivalent to the uncentered model in most situations. In fact, if the grand mean centered model and the uncentered model are estimated with the ML estimator the log-likelihood value for the two models will be identical. Because of that we will not include the grand mean centering in most of the simulation studies, however, it is safe to assume that any conclusions we make about the uncentered model also apply to the grand mean centering model. For completeness we provide the full model,

$$Y_{ij} = \alpha + \beta_2 \overline{X}_{.j} + \beta_{1j}(X_{ij} - \overline{X}_{..}) + \varepsilon_{b,j} + \varepsilon_{w,ij} \tag{33}$$

$$\beta_{1j} = \beta_1 + \xi_j. \tag{34}$$

The hybrid method

The fifth centering option that we discuss here is the hybrid method. This approach is similar to the uncentered method, however, on the between level instead of using the cluster sample mean we use the true/latent mean obtained from the proper decomposition. The model can be expressed as follows:

$$X_{ij} = X_{w,ij} + X_{b,j} \tag{35}$$

$$Y_{ij} = \alpha + \beta_2 X_{b,j} + \beta_{1j} X_{ij} + \varepsilon_{b,j} + \varepsilon_{w,ij} \tag{36}$$

$$\beta_{1j} = \beta_1 + \xi_j \tag{37}$$

where $X_{w,ij}$, $\varepsilon_{b,j}$ and $\varepsilon_{w,ij}$ are assumed to be zero mean normally distributed residuals and $X_{b,j}$ is a normally distribute random effect. This approach has an advantage over the uncentered method because it accounts for the measurement error in $X_{b,j}$. It also has a computational advantage over the latent centering method as it does not involve the product of the two latent variables β_{1j} and $X_{w,ij}$, see equation (22), and, thus, can be estimated with the ML estimator. The method has traditionally been used with the Mplus software prior to the introduction of the latent centering method in Mplus version 8.1. Note also that if the random slope is replaced by a fixed slope, this model becomes an equivalent reparameterization of the latent centering model. However, this model has the same problems as the uncentered model. The model does not separate the within and between effects clearly and lets the random part of the within level slope also be used on the between level. Note also that when the cluster sample size increases, the hybrid model will become identical to the uncentered model (not just a reparameterization but actually identical) because the difference between $X_{b,j}$ and $\overline{X}_{.j}$ will be negligible.

Further discussion

Note that when the covariate X_{ij} is an exogenous variable, i.e., it is not really caused by or formed by the clustering, the latent centering decomposition still provides the most meaningful separation of group level effects and cluster specific effects. Consider, for example, the exogenous variable of employee's gender where employees are nested within companies. The interpretation of $X_{b,j}$ as a predictor would be the percentage of female employees in the company, while the interpretation of $X_{w,ij}$ would be as usual, i.e., it will be a binary variable centered at a particular point. Using the latent centering model given in equations (20–25) would separate the group level and the individual level dynamics. In the case of categorical predictor, however, the latent centering can be implemented on the underlying latent scale. We discuss this in more details in Section 4.4.

Note also that if the covariate X_{ij} has a very small intraclass correlation

$$ICC = \frac{Var(X_{b,j})}{Var(X_{w,ij}) + Var(X_{b,j})} \tag{38}$$

all the centering options become equivalent. Suppose that $Var(X_{b,j}) = 0$. In that case the coefficient β_2 is not identified and has to be removed from the model. This also means that both cluster centering options, the observed and the latent, amount to subtracting a constant from the covariate, i.e., are equivalent to the grand mean centering, which in turn is equivalent to the uncentered and hybrid models. Because of that, the most appropriate and straight forward approach is to simply use the uncentered covariate. Thus, the comparison between the different centering options only applies when the ICC is not zero. It is generally accepted that if the ICC is not statistically significant from 0 the covariate can be treated as uncentered and the effect of $X_{b,j}$ would be marginal can be removed from the model. Note, however, that the statistical significance of the ICC is based on evaluating the hypothesis $v = Var(X_{b,j}) = 0$, which is a boundary hypothesis testing and, thus, is somewhat challenging. For example, with the Bayesian estimation the 95% credibility interval of a variance parameter by default never contains the zero value. An approximate way to do this testing is to use maximum likelihood style t-test using the standard errors reported from the Bayesian estimator. The logic behind this approach is justified by the fact that asymptotically as the number of clusters increases the Bayes estimates and their standard errors are equivalent to the ML estimates and, thus, using a simple t-value approach is not unfounded especially when the number of clusters is large. One can evaluate $t = v/se(v)$ and consider that significant if it is greater than 1.96. This is not a precise test for the variance, however, and should still be considered a rough approximation. A more advanced approach for testing variance significance has been described in Verhagen and

Fox (2012) based on Bayes factor methodology. This method is also implemented in Mplus.

An alternative method for evaluating whether or not a variable can remain uncentered and its cluster level contribution be considered insignificant is to evaluate its design effect, see Muthén and Satorra (1995).

$$DEFF = 1 + (c - 1)ICC \tag{39}$$

where c is the average cluster size. In fact, a commonly used rule of thumb is that a design effect smaller than 2 can be used as a justification for not accounting for the cluster effect. This rule of thumb, however, has recently been shown to be improper in certain situations, see Lai and Kwok (2015).

Note also that another disadvantage of using the latent centering method for a covariate with small ICC is that the model becomes less parsimonious, the estimated coefficients would have wider confidence intervals and the convergence of the estimation will be slower. If the sample size is large, however, this disadvantage would be marginal and, thus, it would still be meaningful to pursue the latent covariate approach even if the ICC is small.

Next, we briefly summarize the availability of estimation methods in Mplus for the various centering options. The Bayesian estimation can be used with all five centering option, with the exception of the hybrid method for a predictor with random slope when the predictor is categorical or it has missing data. The ML estimation can be used with the observed centering, grand mean centering, the uncentered method and the hybrid centering. However, if the predictor has missing data and a random slope the estimation requires numerical integration and becomes impractical for larger models. The ML estimation is also not available for the hybrid method when the predictor is a categorical variable. The ML estimation can be used with the latent centering if the predictor has a non-random slope, however, if the predictor has a random slope the estimation requires a special setup and numerical integration which makes it impractical for larger models. This special setup is described in details in Preacher et al. (2016) in the context of moderation modeling. Note also that the Bayesian estimation in Mplus currently cannot be used for moderation modeling with latent centering and, thus, the ML estimation described in Preacher et al. (2016) is the only available approach. The WLS estimation method can be used with all centering options, except for the hybrid method, when the predictor has a non-random slope.

The Bayesian estimation of the latent centering model is similar to the Bayesian estimation of the two-level SEM model described in Asparouhov and Muthén (2010). The only modification that is needed is as follows. In the standard two-level SEM estimation all between level random effects are updated in one block and the posterior distribution for these random effects is a multivariate normal distribution. In the latent centering model estimation, we split the between level random effects in two blocks. All random effects used for latent centering, such as $Y_{b,j}$ and $X_{b,j}$ in equations (20) and (21), are updated in one block, while all other random effects such as β_{1j} in equation (22) are updated in a separate block. This way we avoid the product of latent variables that causes difficulties in the ML estimation. The posterior distribution for each of the two blocks is again a multivariate normal distributions because it is conditioned on the other block. No further modifications are needed in the MCMC estimation.

Throughout this article the priors used with the Bayesian estimation are the Mplus default priors. These priors are uninformative or weakly informative and can be found in Asparouhov and Muthén (2010).

The simulation studies presented here are based on the assumption that the observations in the sample belonging to the same cluster represent a small random sample from that cluster. When we generate data from a hypothesized model, we always use the true cluster mean $X_{b,j}$ in the model instead of the sample mean $\overline{X}_{.j}$. If $\overline{X}_{.j}$ is used to generate the data for Y_{ij} we would create an illogical loop where the population distribution is affected by the random sample selected from that population. The sampling scheme should not have an effect on the target population model. In a logical simulation study a large target population is generated first which is then sampled from and this is incompatible with using $\overline{X}_{.j}$ in the model. If $\overline{X}_{.j}$ is used in the model instead of $X_{b,j}$, we first have to generate the target population for covariate X_{ij} only, then sample from the covariate target population, then form $\overline{X}_{.j}$ and then use that to generate the dependent variable Y_{ij} in the target population. This of course creates the illogical loop where the value of one member of the cluster depends on which other observations were sampled from that cluster, i.e., the sampling scheme becomes an integral part of the model and the generation of the target population. This would be very difficult to justify. Because we always use the true mean for the data generation purposes rather than the sample mean, what appears as a favoritism on our part toward the latent centering methodology is actually not so at all. It is rather driven by fundamental statistical principles.

In the next sections, we illustrate the advantages of the latent centering model through simulations studies. All Mplus inputs and outputs for these simulation studies can be found at statmodel.com.

THE TWO-LEVEL REGRESSION MODEL

Let Y_{ij} and X_{ij} be the dependent and the predictor variables. Consider the following two-level regression model.

$$X_{ij} = X_{w,ij} + X_{b,j} \tag{40}$$

$$Y_{ij} = \alpha_j + \beta_j X_{w,ij} + \varepsilon_{ij} \tag{41}$$

$$\varepsilon_{ij} \sim N(0, \sigma), X_{w,ij} \sim N(0, \psi) \tag{42}$$

$$\begin{pmatrix} X_{b,j} \\ \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \right) \tag{43}$$

We generate 100 samples according to the above model using 1000 clusters of size 15. We use a large number of clusters in this simulation so that any biases in the parameter estimates can be seen clearly. The data is analyzed with the latent centering method using Bayesian estimation, the observed group mean centering method using ML estimation, the uncentered method using ML estimation and the hybrid method using ML estimation.

The results are reported in Table 1, which also contains the true parameter values used for the data generation. In these data the ICC for the covariate X_{ij} is 0.5 while for the dependent variable Y_{ij} is 0.3. The latent centering method yields consistent estimates with confidence interval coverage near the nominal levels. The observed group mean centering method yields consistent estimates for the means of the random effects but some of the variance covariance parameters are biased and the coverage drops below the nominal levels. These biases could result in structural biases if we are to estimate structural or regression models on the between level involving these random effects such as for example the regression model in equation (13). The uncentered method yields poor results. This is mostly due to the fact that the model has to be reparameterized before it can be used to make implications for the generating model (44–43). Such reparameterizations are fairly simple when it comes to the μ_i parameters but will become increasingly difficult for random effects variance covariance parameters. The reparameterization would also become increasingly difficult when the model's complexity increases, for example, when having two predictor variables. Consider for example the parameter $\mu_2 = E(\alpha_j)$. Because the model is uncentered the proper reparameterization implies that the random intercept in this model is not just α_j but it is $\alpha_j + \beta_j \overline{X}_{.j}$ with expected value $\mu_2 + \mu_3\mu_1 + \sigma_{13}$. The average estimates in our simulations using the uncentered method are $\mu_2 = .46, \mu_1 = \mu_3 = 1, \sigma_{13} = 0.54$. Thus, if we compute the expected value for the random intercept in this case we get $0.46 + 1 + 0.54 = 2$ which is the true value. Because the uncentered method does not seem to have any benefits, however, it is difficult to recommend it. The observed group mean centering method's bias is fairly small and will be negligible when the cluster sizes increase to 100 or larger. When the cluster sample sizes are smaller, however, the latent centering method clearly has the advantage. The hybrid uncentered method also does not perform well. As expected the method produced estimates close to the uncentered method. It eliminates the biases of the uncentered method for the two parameters σ_{11} and σ. However, as for the uncentered method, a complex model transformation is needed to obtain the results in the original latent centered metric.

TABLE 1
The Two-Level Regression Model: Absolute Bias(Coverage)

Parameter	*True Value*	*Latent*	*Observed*	*Uncentered*	*Hybrid*
μ_1	1	0.00 (0.93)	0.00 (0.95)	0.00 (0.95)	0.00 (0.95)
μ_2	2	0.00 (0.93)	0.00 (0.96)	1.46 (0.00)	1.46 (0.00)
μ_3	1	0.00 (0.95)	0.00 (0.95)	0.00 (0.96)	0.00 (.96)
σ_{11}	1	0.01 (0.93)	0.07 (0.73)	0.07 (0.73)	0.00 (0.94)
σ_{22}	1	0.01 (0.91)	0.13 (0.32)	1.75 (0.00)	1.75 (0.00)
σ_{33}	1	0.01 (0.93)	0.00 (0.93)	0.10 (0.39)	0.10 (0.39)
σ_{12}	0.5	0.00 (0.93)	0.06 (0.63)	1.45 (0.00)	1.45 (0.00)
σ_{13}	0.5	0.01 (0.94)	0.00 (0.96)	0.04 (0.78)	0.04 (0.78)
σ_{23}	0.5	0.01 (0.91)	0.00 (0.93)	1.38 (0.00)	1.38 (0.00)
σ	1	0.00 (0.96)	0.00 (0.97)	1.00 (0.00)	0.01 (0.94)
ψ	1	0.00 (0.96)	0.07 (0.00)	0.01 (0.94)	0.00 (0.97)

Next we consider a two-level regression model with multiple covariates. Let Y_{ij} be the dependent variable and X_{ijk} be the $k-$ th predictor variable, $k = 1, ..., K$. Consider the following regression model:

$$X_{ijk} = X_{w,ijk} + X_{b,jk} \tag{44}$$

$$Y_{ij} = \alpha_j + \sum_{k=1}^{K} \beta_{jk} X_{w,ijk} + \varepsilon_{ij} \tag{45}$$

$$\varepsilon_{ij} \sim N(0, \sigma), X_{w,ijk} \sim N(0, \psi_k). \tag{46}$$

The model has $2K + 1$ between level random effects: α_j, β_{jk} and $X_{b,jk}$ are assumed to have a multivariate normal distribution. We conduct a simulation study with $K = 4$ covariates. The cluster size is set to 15 as in the previous simulation study. The number of clusters in each sample is set to either $N = 100$, $N = 300$ or $N = 500$. Note that the case with 100 clusters should be viewed as small sample size estimation. The within level model for Y_{ij} has 5 cluster specific parameters to be estimated with 15 observations in the cluster. The ratio of the number of observations to the number of parameters on the within level is thus 3. On the between level, if we estimate only the means and the variance covariance of the intercept and slope random effects of Y_{ij}, that ration would be 5. If we estimate the means and the variance covariance of all 9 random effects the ratio is approximately 2. A simple rule of thumb is that when this ratio is smaller than 5, asymptotic theory cannot be relied upon and special small sample size considerations apply. The parameters used for the data generation are as follows: $\sigma = \psi_k = 1$ and the correlation between the covariates on the within level $X_{w,ijk}$ is set to 0.3. The mean of the random intercept α_j is set to 0, the mean of the random slopes β_{jk} is set to 1 and the means of the random intercepts for the covariates $X_{b,jk}$ is set to 0. The variance covariance matrix for the 9 random effects on the

TABLE 2
The Two-Level Regression Model with Multiple Covariates: Absolute bias/coverage/MSE for $Var(\alpha_j)$

Centering	*Estimator*	*Covariates*	$N = 500$	$N = 300$	$N = 100$
Latent	Bayes	Endogenous	0.09/0.88/0.02	0.14/0.87/0.04	0.40/0.72/0.25
Latent	Bayes	Exogenous	0.07/0.83/0.02	0.10/0.86/0.03	0.23/0.85/0.12
Observed	ML	Endogenous	0.90/0.00/0.84	0.91/0.00/0.87	0.89/0.06/0.91
Observed	ML	Exogenous	0.90/0.00/0.84	0.91/0.00/0.87	0.89/0.06/0.91
Observed	Bayes	Endogenous	0.97/0.00/0.98	1.05/0.00/1.15	1.35/0.00/2.00
Observed	Bayes	Exogenous	0.96/0.00/0.96	0.97/0.00/0.99	1.18/0.01/1.54

between level is set to the matrix with 1 on the main diagonal and 0.5 for all off diagonal entries.

There are two ways to estimate the above model with respect to how the covariates are treated: the covariates can be treated as exogenous or as endogenous variables. If the covariates are treated as exogenous variables the focus is on estimating the conditional distribution of $[Y|X]$ and in this case the covariance between $X_{w,ijk}$ and the covariances between $X_{b,jk}$ and the other random effects are not included in the model. If the covariates are treated as endogenous variables these covariances are all included. The exogenous treatment has the advantage that it yields a more parsimonious model which can be very beneficial in small sample size situations. The endogenous model has the advantage that it provides a more comprehensive model which represents the data better and can be more accurate for large samples due to including more data in the model estimation (e.g., using the between level means for the covariates as covariates for the random effects). We consider again the latent centering approach based on the Bayesian estimation. We also consider the observed centering approach using the ML and the Bayesian estimation. Because of the small sample size we can expect some difference between the two estimators. The quality of the estimation will be evaluated by considering the estimation results for the variance of the random intercept parameter which was among the most problematic parameter in Table 1.

The results of this simulation study based on 100 replications are presented in Table 2. The latent centering method outperforms the observed centering substantially in all cases. The latent centering with exogenous covariates outperformed the latent centering with endogenous covariates when $N = 100$ and, thus, we see the exogeneity benefit of parsimony exists with small samples in these circumstances. For the larger samples, however, the endogenous treatment of the covariates performs equally well and in fact outperformed the exogenous treatment for other model parameters, not presented here. Thus, for larger samples, the endogenous treatment should be preferred. The results in Table 2 also show that the problems with the observed centering exist regardless of the estimator. We can also see that the ML estimator is not affected by the model choice for the covariates. That is due to the fact that the log-likelihood of the model with observed centering can be represented as the sum of two independent log-likelihood terms: one for $[Y|X]$ and one for $[X]$, i.e., the inclusion of X in the model does not affect the estimation of $[Y|X]$ in any way. This does not seem to be the case when using the Bayes estimator and the covariate treatment affects the results.

LÜDTKE'S BIAS

In multilevel models the contextual effect occurs when the aggregate predictor $\overline{X}_{.j}$ affects the outcome, even after controlling for the individual level predictor X_{ij}, see Raudenbush and Bryk (2002). In the presence of contextual effects, Lüdtke et al. (2008) shows that the observed mean centering method yields biased results, while the latent mean centering method is unbiased. In the following sections, we illustrate Lüdtke's bias in the case of non-random/fixed slope and then we extend that to the case of multilevel regression with random slope.

Fixed slope

Consider the following two-level regression model,

$$X_{ij} = X_{w,ij} + X_{b,j} \tag{47}$$

$$Y_{ij} = \alpha_j + \beta_1 X_{w,ij} + \varepsilon_{w,ij} \tag{48}$$

$$\alpha_j = \alpha + \beta_2 X_{b,j} + \varepsilon_{b,j} \tag{49}$$

$$\varepsilon_{w,ij} \sim N(0, \sigma_w), \varepsilon_{b,j} \sim N(0, \sigma_b), X_{w,ij} \sim N(0, \psi_w), X_{b,j} \sim N(\mu, \psi_b). \tag{50}$$

There are no random slopes in this model. The within and the between part of the the covariate X affects the dependent variable Y but the effect for each of these two components are different. The contextual effect is defined as $\beta_2 - \beta_1$. This model has been studied extensively in Asparouhov and Muthén (2006a) and Lüdtke et al. (2008) in the context of comparing the latent mean centering and the observed group mean centering methods. It was found that the observed group mean centering method yields a bias in the key coefficient β_2. This bias can be computed explicitly in terms of the model parameters and is approximately equal to

$$\frac{(\beta_1 - \beta_2)\psi_w/n}{\psi_b + \psi_w/n}, \tag{51}$$

where n is the size of the clusters. We refer to this bias as Lüdtke's bias. The bias can also be expressed in terms of the reliability of the sample cluster mean for X_{ij}

$$R = \frac{Var(X_{b,j})}{Var(\overline{X}_{.j})} = \frac{\psi_b}{\psi_b + \psi_w/n} = \frac{n \cdot ICC}{(n-1)\cdot ICC + 1} \tag{52}$$

where ICC is the intra-class correlation given in (38). Thus, an equivalent expression for Lüdtke's bias is

$$(\beta_1 - \beta_2)\frac{(1-ICC)/n}{ICC + (1-ICC)/n} = (\beta_1 - \beta_2)(1-R). \tag{53}$$

Next we illustrate this bias with the following simulation study. We generate 100 samples with 500 clusters of size 15 and analyze the data using the observed and the latent centering methods. Because there are no random slopes in this model both models can be estimated with the ML estimator. The results are presented in Table 3, as well as the true parameter values used for the data generation. Note that we used regression coefficients that are in opposite direction. This is known as the big fish small pond (BFSP) effect and is common in education studies, see Marsh et al. (2009) and Marsh et al. (2012). The simulation results show that the bias in the observed group mean centering method is not limited to the parameter β_2 but can be seen in several other parameters, including variance parameters. If the model is expanded further to include structural equations for the random effects these additional biases can lead to biases in structural parameters. Notably, however, the within level regression parameter β_1 is unbiased. The latent centering method provides unbiased estimates and coverage near the nominal level and, thus, should be preferred to the observed group mean centering method. As the cluster sample sizes are increased to larger values the difference between the latent and the observed group mean centering method will become negligible. However, if the ICC of the covariate decreases, Lüdtke's bias increases. Note also that in the above model the uncentered method is equivalent to the observed group mean centered method.

TABLE 3
Lüdtke's Bias: Absolute Bias(Coverage)

Parameter	*True Value*	*Latent Centering*	*Observed Centering*
α	2	0.01 (0.95)	0.15 (0.46)
β_1	−1	0.00 (0.96)	0.00 (0.96)
β_2	1	0.00 (0.98)	0.14 (0.17)
σ_w	1	0.00 (0.95)	0.00 (0.95)
σ_b	0.9	0.01 (0.94)	0.24 (0.10)
μ	1	0.00 (0.91)	0.00 (0.91)
ψ_w	1	0.00 (0.95)	0.06 (0.01)
ψ_b	0.9	0.01 (0.97)	0.06 (0.90)

Random slope

In this section, we conduct a simulation study for the case when the regression coefficient on the within level is a random effect. This is a natural extension of the model described in the previous section as the effect of the covariate can be expected to vary across clusters. Consider the model

$$X_{ij} = X_{w,ij} + X_{b,j} \tag{54}$$

$$Y_{ij} = \alpha_j + \beta_{1,j}X_{w,ij} + \varepsilon_{w,ij} \tag{55}$$

$$\alpha_j = \alpha + \beta_2 X_{b,j} + \varepsilon_{b,j} \tag{56}$$

$$\beta_{1,j} = \beta_1 + \beta_3 X_{b,j} + \xi_{b,j} \tag{57}$$

$$\begin{pmatrix}\varepsilon_{b,j}\\ \xi_{b,j}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}\right) \tag{58}$$

$$\varepsilon_{w,ij} \sim N(0,\sigma_w), X_{w,ij} \sim N(0,\psi_w), X_{b,j} \sim N(\mu,\psi_b). \tag{59}$$

In this model the between part of the covariate $X_{b,j}$ is used as a predictor not just for the random intercept α_j but also for the random slope $\beta_{1,j}$. We generate 100 samples with 500 clusters of size 15 using the above model and we analyze the data using the latent and observed group mean centering. The latent centering model can be estimated only with the Bayesian estimator. We estimate the observed group mean centering model with the ML estimator. The results of the simulation are presented in Table 4. The latent centering method appears to work well, while the observed group mean centering method shows bias not just for the β_2 coefficient but also, for example, for the within level effect $\beta_{1,j}$. Both coefficients in equation (57) have biased estimates.

TABLE 4
Lüdtke's Bias with Random Slope: Absolute Bias(Coverage)

Parameter	*True Value*	*Latent Centering*	*Observed Centering*
α	2	0.00 (0.95)	0.06 (0.85)
β_1	−1	0.01 (0.89)	0.06 (0.84)
β_2	1	0.00 (0.96)	0.06 (0.71)
β_3	1	0.00 (0.97)	0.06 (0.73)
σ_w	1	0.00 (0.97)	0.00 (0.98)
σ_{11}	0.9	0.02 (0.94)	0.20 (0.21)
σ_{12}	0.5	0.01 (0.96)	0.06 (0.87)
σ_{22}	1	0.01 (0.96)	0.06 (0.95)
μ	1	0.01 (0.97)	0.01 (0.98)
ψ_w	1	0.00 (0.98)	0.06 (0.01)
ψ_b	0.9	0.01 (0.94)	0.07 (0.84)

Let us also consider the grand mean centering method as well as the uncentered method for the above model. These two methods are equivalent and a reparameterization of each other, however, unlike the case with a non-random slope, these two methods are not equivalent to the observed group mean centering method. In fact, if all three methods are estimated as conditional models where all covariates are conditioned on and not included in the likelihood, we can use the BIC criterion to compare the models. Table 5 shows the BIC results for the three models. As expected the uncentered model and the observed grand mean centering yield identical results. In addition the results indicate that observed group mean centering model is the better fitting model. The BIC is smaller for the observed group centering in this simulation, not just on average, but for each individual replication. This is as expected because the observed group centering model is closer to the true model. The latent centering model cannot be compared with an information criterion to the models in Table 5 not just because it is based on the Bayesian estimator, which produces the DIC criterion instead of the BIC criterion, but also because the model treats the X_{ij} as a dependent variable. Thus, the likelihood for that model is the joint likelihood for Y_{ij} and X_{ij} rather than the conditional likelihood $[Y_{ij}|X_{ij}]$.

One final aspect that we can consider in this simulation is the precision of the estimates in terms of MSE (mean squared error) or the SMSE (square root of the mean squared error) for the five different centering options. Table 6 contains the results for the β_2 parameter as well as the reparameterization formulas that are needed with the grand mean centering, the uncentered and the hybrid methods. While the grand mean method and the uncentered method yield equivalent models the reparameterization formulas are different. Because of that and the fact that the uncentered method uses a non-linear expression, the standard errors for these two methods are different even though the point estimates are the same. This is the result of using the delta method with the ML estimator. If we estimate the models with the Bayesian estimator the standard errors would be the same for the two equivalent models. The bias for the group-mean centering, the grand mean centering and the uncentered method is about the same, however, the MSE is not, and the observed group mean centering method is much more accurate than the grand mean centering method and the uncentered method. At the same time the latent centering method is more accurate than the observed group mean centering method. The hybrid method appears to have improved the bias and the coverage. However, in terms of MSE it is not any better than the uncentered method. The MSE differences are so large that we can expect to see this effect for other similar models and that this is a general phenomenon. Note here that when the random slope $\beta_{1,j}$ is not random the MSE for the latent centering method is also much smaller that that of the observed group-mean centering, but the observed group-mean centering has the same MSE as the uncentered and the grand-mean centering. We can generally expect that this dramatic increase in MSE for the grand-mean centering and the uncentered method would diminish as the variance of the random slope is closer to 0. Similarly, the hybrid method is equivalent to the latent centering method for non-random slopes. That means that the MSE increase associated with the hybrid method also depends on the size of the variance of the random slope and would diminish as the random slope variance approaches 0.

TABLE 5
Lüdtke's Bias with Random Slope: Average BIC

Centering	*Observed-Group*	*Observed-Grand*	*Uncentered*
BIC	23,965	24,578	24,578

Multilevel probit regressions

In this section, we illustrate with simulation studies Lüdtke's bias in multilevel probit regressions with or without random slopes. As in the linear regression case, the bias can be observed when the cluster sizes are small and there is a contextual effect. Here, we discuss the multilevel probit regression, however, the conclusions also apply to multilevel logistic regression. In the simulation studies we use a binary dependent variable Y_{ij} but the results are similar for a categorical variable with more than two categories. First, we consider the multilevel probit regressions without a random slope.

$$X_{ij} = X_{w,ij} + X_{b,j} \tag{60}$$

TABLE 6
Lüdtke's Bias with Random Slope: Results for $\beta_2 = 1$

Centering	*Latent*	*Observed Group*	*Observed Grand*	*Uncentered*	*Hybrid*
Reparameterization	β_2	β_2	$\beta_1 + \beta_2$	$\beta_1 + \beta_2 + 2\mu\beta_3$	$\beta_1 + \beta_2 + 2\mu\beta_3$
Estimate	1.00	0.94	0.95	0.95	1.02
Standard error	0.054	0.049	0.093	0.122	0.128
Coverage	096	0.71	0.87	0.94	0.96
SMSE	0.054	0.082	0.123	0.123	0.121

$$P(Y_{ij} = 1) = \Phi(\alpha_j + \beta_1 X_{w,ij}) \tag{61}$$

$$Y_{b,j} = \alpha_j = \alpha + \beta_2 X_{b,j} + \varepsilon_{b,j} \tag{62}$$

$$\varepsilon_{b,j} \sim N(0, \sigma_b), X_{w,ij} \sim N(0, \psi_w), X_{b,j} \sim N(\mu, \psi_b). \tag{63}$$

The contextual effect here again exists when β_1 and β_2 are not identical. We conduct three simulation studies with 100 data sets, 500 clusters of size L. For simulations 1 and 2 we use $L = 15$ and for simulation 3 we use $L = 50$ to evaluate the effect of cluster size on the bias. The parameters used for data generation purposes are as follows $\alpha = 0.5$, $\psi_b = \sigma_b = 0.9$, $\mu = 0.4$, $\beta_2 = 1$ and $\psi_w = 1$. We vary the value of β_1 across the three simulation studies to evaluate the effect of the contextual effect on the bias. In simulations 1 and 3, the value of β_1 is -1, which creates a contextual effect, while in simulation 2 it is 1, i.e., no contextual effect. The quality of the estimation will be evaluated using a single parameter $r = Cor(Y_{b,j}, X_{b,j})$. This parameter is not a model parameter but is derived from the estimated model parameters as follows:

$$r = \frac{\beta_2 \psi_b}{\sqrt{\psi_b(\sigma_b + \beta_2^2 \psi_b)}}. \tag{64}$$

We analyze the data using the latent centering method with the Bayes estimator and the observed centering method with the ML estimator. In this situation, the estimator choice does not matter. The Bayes and the ML estimators are asymptotically identical and because the number of clusters is 500 in all three simulations the choice of the estimator is irrelevant. Both the observed and the latent centering can be used with either estimator. The ML estimation with the observed centering uses a one dimensional numerical integration. The results of the simulation are presented in Table 7. The latent centering approach performs well while the observed centering shows poor results in the presence of contextual effect. The bias in the observed centering method decreases as the cluster size increases. In addition, formula (53) predicts the bias of the contextual effect in the multilevel probit regressions as well.

Consider now a multilevel probit regression with random slope. Equations (60), (62) and (63) remain the same while equation (61) is replaced by

$$P(Y_{ij} = 1) = \Phi(\alpha_j + \beta_{1j} X_{w,ij}). \tag{65}$$

where $\beta_{1j} \sim N(\beta_1, v)$ is a cluster specific random effect. We generate data using the same parameter values as in the previous simulation and the value of v is set to 0.5. We analyze the data using observed centering with the ML estimator based on a two dimensional numerical integration and with the latent centering method using the Bayes estimation which was introduced in Mplus V8.1. The ML estimation cannot be used to estimate the model with latent centering directly but an approximation model can be estimated with an additional dimension of integration. We do not consider this further, however, as such an estimation is computationally intensive. The results of the simulation studies are presented in Table 8. The latent centering approach works very well in this context as well, while the observed centering approach shows similar problems to the problems observed for the multilevel probit regression without random slopes.

TABLE 7
Lüdtke's Bias in Multilevel Probit Regression: Absolute Bias (Coverage) for r

Cluster Size	*Contextual Effect*	*Latent Centering*	*Observed Centering*
15	Yes	0.00 (0.95)	0.09 (0.08)
15	No	0.00 (0.92)	0.01 (0.95)
50	Yes	0.00 (0.93)	0.03 (0.81)

TABLE 8
Lüdtke's Bias in Multilevel Probit Regression with Random Slope: Absolute Bias(Coverage) for r

Cluster Size	*Contextual Effect*	*Latent Centering*	*Observed Centering*
15	Yes	0.01 (0.93)	0.09 (0.09)
15	No	0.00 (0.96)	0.00 (0.96)
50	Yes	0.00 (0.94)	0.03 (0.84)

Multilevel regression with a categorical predictor or mediator

In the previous section we considered the multilevel model where a categorical variable is predicted by a continuous variable. In this section, we reverse the roles of the two variables, i.e., the continuous variable is predicted by a categorical variable. For simplicity, we use a binary categorical variable but the discussion applies to ordered categorical variables with more than two categories. Denote the binary variable by X_{ij} and the continuous dependent variable by Y_{ij}. The most common modeling approach in this situation consists of not centering the binary predictor at all. This approach, however, does not allow us to estimate a contextual effect. Enders and Tofighi (2007) proposed to treat the binary variable as continuous and use the observed group centering. In this section, we compare the latent centering approach to the observed centering approach. We show that Lüdtke's bias occurs in this situation as well when using observed centering and that this bias can be resolved with latent centering. Consider the following two-level model:

$$Y_{ij} = Y_{w,ij} + Y_{b,j} \tag{66}$$

$$X^*_{ij} = X^*_{w,ij} + X^*_{b,j} \tag{67}$$

$$Y_{w,ij} = \beta_1 X^*_{w,ij} + \varepsilon_{w,ij} \quad (68)$$

$$Y_{b,j} = \alpha_j = \alpha + \beta_2 X^*_{b,j} + \varepsilon_{b,j} \quad (69)$$

$$X_{ij} = 0 \Leftarrow X^*_{ij} < \tau \quad (70)$$

$$P(X_{ij} = 0|j) = \Phi(\tau - X^*_{b,j}) \quad (71)$$

$$\varepsilon_{b,j} \sim N(0, \sigma_b), \varepsilon_{w,ij} \sim N(0, \sigma_w), X^*_{w,ij} \sim N(0, 1), X^*_{b,j} \sim N(0, \psi_b). \quad (72)$$

This model represents a standard two-level model where the structural relationships occur on the underlying latent scale between the continuous and the categorical variable, i.e., it is based on estimating a polyserial correlation between the variables rather than the Pearson type correlation on the observed scale. The variable X^*_{ij} represents the underlying continuous variable that is being categorized to obtain the observed binary variable. This variable is further decomposed as a within-between and the two components are denoted by $X^*_{w,ij}$ and $X^*_{b,j}$. The variable $X^*_{b,j}$ represents the cluster specific deviation of the threshold parameter τ while the within component is the centered zero mean effect. As usual the variance of $X^*_{w,ij}$ is fixed to 1, similarly to how this is done in a standard two-level probit regression. The variable $X^*_{b,j}$ also represents the mean of X^*_{ij} in cluster j and, therefore, the regression on the within level (68) represents the latent centered regression.

We generate data according to the above model and analyze the data by estimating the same model using the Bayes estimator. Because the binary variable is modeled explicitly, missing at random data can be accommodated. The missing data is imputed within the Bayesian estimation from the above multilevel polyserial correlation model conditional on the observed data. In addition, the binary variable can be further regressed on another variable resulting in a two-level probit regression as in the previous section. Therefore, we can easily extend the above model to a two-level mediation model with a binary mediator. Both of these features are not possible with the observed centering. In Mplus the above two-level model can also be estimated with the WLS family of estimators, however, such estimation does not accommodate MAR missing data as well as random slopes in the within level regression (68). We also estimate the above model using observed centering as suggested by Enders and Tofighi (2007). Because the dependent variable is continuous such estimation can be done with the ML estimator without numerical integration. Note here that with the observed centering, the regression on the within and the between level is done on the observed scale. To be able to properly evaluate the ability of the methods to estimate well the contextual effect we will compute the between level correlation $r = Cor(Y_{b,j}, X_{b,j})$ as was done in the previous section. This will eliminate the scale differences between the models while focusing on the strength of the relationship of the variables on the between level. To generate the data we use the following parameters $\alpha = 0.4, \beta_2 = 1, \tau = -0.5, \sigma_w = 1, \sigma_b = \psi_b = 0.9$. The parameter β_1 will take two values 1 or -1. The value of -1 creates a contextual effect while the value of 1 does not. We generate 100 data sets with 500 clusters and cluster sizes 15, 50 and 100 with and without contextual effect.

TABLE 9
Lüdtke's Bias in Multilevel Regression with Binary Predictor: Absolute Bias(Coverage) for r

Cluster Size	*Contextual Effect*	*Latent Centering*	*Observed Centering*
15	Yes	0.00 (0.94)	0.12 (0.01)
50	Yes	0.00 (0.95)	0.06 (0.35)
100	Yes	0.00 (0.93)	0.04 (0.63)
15	No	0.00 (0.91)	0.03 (0.79)
50	No	0.00 (0.95)	0.03 (0.87)
100	No	0.00 (0.93)	0.03 (0.80)

The results are presented in Table 9. Lüdtke's bias can be seen here as well when the model is estimated with the observed centering in the presence of contextual effect. The latent centering method produces no bias and the coverage is near the nominal levels. We can also see in Table 9 that the bias in the observed centering estimation does not disappear completely when the data is generated without the contextual effect or when the cluster size increases. This suggests that there is an additional measurement error that has not been observed previously. The observed centering uses on the between level the variable $p_j = E(X_{ij}|j) = \Phi(X^*_{b,j} - \tau)$ instead of $X^*_{b,j}$. There are two problems with that. First, the function Φ is not a linear function and this will distort the estimate of the correlation r. Second, the transformation Φ essentially truncates the distribution of $X^*_{b,j}$ at 0 and 1. It is well known that truncation attenuates correlation parameters, see Muthén (1990). This additional bias depends mainly on the ICC of the binary variable. The bigger the ICC is, the bigger the range in the probability values p_j across clusters. The bigger that range is, the bigger the discrepancy between Φ and a linear function is within that range and, therefore, the bigger the bias. The ICC for the categorical predictor in the above simulation studies is 0.47. For ICC of 0.25 or lower, however, we can expect that this additional bias will be negligible. In summary, we see that using the observed centering for a categorical predictor causes an additional bias in the contextual effect due to the nonlinearity of the link function Φ.

There is bigger problem, however, with the observed centering for categorical predictors. When using the observed centering, the within level correlation estimate is substantially biased, while the latent centering yields unbiased estimates. In the above simulations the observed centering produces (by absolute value) a within level correlation of approximately 0.46 while the true value is 0.71. This discrepancy is identical across all simulation settings, i.e., with and without contextual effect, and with small or large cluster sizes. Note that this level of attenuation is substantially larger than what is typically observed when the

Pearson correlation is used instead of the polyserial correlation. One plausible explanation for this is that there are further model violations. One such model violation is the fact that the variance of the observed centered within part of the categorical variable is not constant across clusters. That variance is $p_j(1-p_j)$, i.e., not only it is not constant but it is directly determined by the between part of the variable that we are using for the centering. This exposes another severe model violation. The within and the between parts of the categorical variable are not independent variables. The between part of the variable determines completely the variance of the within part and also its range. For example, when the between part is 0.3, the possible range for the within part are only the two values −0.3 and 0.7. This dependence between the within and the between level components of the binary variable severely undermines the concept of contextual effect. We are no longer able to estimate the effect of the cluster component as a separate effect. Some of the cluster level effect will still be channeled through the within level regression. Yet another model violation is the normality assumptions for the within and the between components. In fact, the binary distribution on the within level can be severely skewed. Without any doubt, however, the biggest contributor to the bias on the within level is the fact that the observed values used with the observed centering are simply measurements of the underlying continuous concept. These measurements have a substantial measurement error that is not accounted for.

Let us summarize our findings. The observed centering for the binary variable is difficult to recommend because of the numerous problems we described above. The real question is when we should use the above latent centering model instead of using the uncentered method with the binary predictor used on the observed 0/1 scale. No doubt the estimation of the uncentered model is simpler and this is definitely a reason to prefer the uncentered model. In fact that has been the main reason such models have been preferred in the past. With the advances in the Bayesian methodology, however, this is no longer a compelling reason. Enders and Tofighi (2007) argue that the uncentered approach may actually be the right approach to address a particular substantive question. Another argument that often is invoked against the latent centering model is the fact that it uses an underlying concept. In certain situations, such a concept is indeed questionable. For example, if the binary predictor is gender, it is difficult to argue that there is an underlying construct that is dichotomized to yield the observed values. We agree that this is indeed the case for certain situations, however, here we will list several reasons for preferring the latent centering approach. If the distribution of the binary predictor is not the same across the clusters, the variable is no longer a pure predictor as it is affected by another variable, i.e., the clustering. It is better to model that cluster specific distribution as in the latent centering approach through a multilevel probit model. Failing to do so can result in model misspecifications. Failing to establish a cluster invariant distribution for the binary predictor can result in model misspecifications. If the ICC of the binary predictor is substantively and statistically significant, the distribution of the variable is cluster specific and should be modeled as such. In addition, there may indeed be a contextual effect and that is best modeled through the latent centering approach as it resolves Lüdtke's bias. Also, suppose that the binary predictor is not just a predictor but is also a mediator variable that itself is regressed on another covariate. This situation is best handled through the latent centering approach and the multilevel probit regression with a random intercept and possibly a random slope for the additional covariate. Another reason to prefer the latent centering approach is the case when there is missing data on the binary predictor. Proper handling of the missing data requires modeling the joint distribution of $[Y,X]$ instead of the conditional distribution of $[Y|X]$. Thus, the two-level probit regression with random intercept for X and the latent centering is needed here as well. In single level settings, Mplus implements a special methodology using the ANALYSIS option PREDICTOR = OBSERVED, see Section 7 in Asparouhov and Muthén (2010). This methodology allows us to model $[Y,X]$ where the regression of Y on X is performed on the observed scale even when there is missing data on X. That methodology, however, is currently not available in two-level settings and the best approach remains the latent centering model where X^* is used as the predictor for Y. Further discussion of categorical mediators can be found in section 8.4 in Muthén, Muthén, and Asparouhov (2017).

Next we consider the situation where the multilevel regression of Y on X has a random slope. In Mplus 8.1 it is possible to estimate this model using the latent centering approach with the Bayes estimator. The model is the same as in equations (66–72), where equation (68) is now replaced by:

$$Y_{w,ij} = \beta_{1j} X^*_{w,ij} + \varepsilon_{w,ij} \tag{73}$$

$$\beta_{1j} \sim N(\beta_1, v). \tag{74}$$

In this simulation, we generate data according to the above model and we analyze the data using latent and observed

TABLE 10
Lüdtke's Bias in Multilevel Regression with Binary Predictor and Random Slope: Absolute Bias(Coverage) for r

Cluster Size	*Contextual Effect*	*Latent Centering*	*Observed Centering*
15	Yes	0.01 (0.99)	0.07 (0.65)
50	Yes	0.00 (0.96)	0.05 (0.94)
15	No	0.01 (1.00)	0.01 (0.98)
50	No	0.00 (0.96)	0.00 (0.93)

centering. To generate the data we use the following parameters $\alpha = 0.4$, $\beta_2 = 1$, $\tau = 0$, $\sigma_w = 1$, $\sigma_b = 0.9$, $\psi_b = 0.4$ and $v = 0.5$. The parameter β_1 will take two values 1 or -1 to evaluate the impact of contextual effect on the estimation. We generate 100 data sets with 100 clusters and cluster sizes 15 and 50, with and without contextual effect. The results are presented in Table 10. We can see from these results that Lüdtke's bias can be seen here as well for the observed centering approach in the presence of contextual effect. We can also see that the latent centering approach resolves that bias. Because we used a lower value for the ICC of the categorical predictor, the bias due to the non-linearity of the link function is not present in this simulation study.

Note that the above model allows us to estimate cluster specific polyserial correlations through the random slopes. Thus, the polyserial correlation is essentially modeled itself as a random effect. This is a unique feature of the models discussed in this and the previous section. The latent centering methodology also extends to the case where both variables, the predictor and the dependent variable, are categorical. Lüdtke's bias can be found in that case as well and can be resolved again with the latent centering.

LATENT CENTERING IN TIME-SERIES MODELS

In time-series models there are two separate centering issues. First, we have the contemporaneous centering where we center a predictor from the same time period. This as usual is subject to Lüdtke's bias if the observed centering is used. However, additional biases can occur with the observed centering when the autocorrelation is ignored. Second, we have the lag centering, i.e., the centering of the lag copy of the dependent variable used as a predictor in an autoregressive model. This situation is subject to Nickell's bias if the observed centering is used. The bias occurs in the mean of the random autocorrelation. It is shown in Asparouhov et al. (2018) that latent centering can be used to resolve Nickell's bias. In this section we focus on the biases that occur with the contemporaneous observed centering due to the autocorrelation in the data and on the lag centering in the context of categorical variables and the random tetrachoric autocorrelation. We show how the latent centering can be used to resolve the bias in both of these situations.

Contemporaneous centering: Lüdtke's bias with autocorrelation

Let Y_{it} and X_{it} be the dependent variable and the covariate for individual i at time t. The two-level time-series regression model is described as follows:

$$X_{it} = X_{b,i} + X_{w,it} \tag{75}$$

$$Y_{it} = \alpha_i + \beta_1 X_{w,it} + \varepsilon_{it} \tag{76}$$

$$\alpha_i = \alpha + \beta_2 X_{b,i} + \varepsilon_i \tag{77}$$

$$\varepsilon_{it} = r_y\, \varepsilon_{i,t-1} + \delta_{it} \tag{78}$$

$$X_{w,it} = r_x X_{w,i,t-1} + \xi_{it} \tag{79}$$

$$\delta_{it} \sim N(0, \sigma_1), \xi_{it} \sim N(0, \psi_1), \varepsilon_i \sim N(0, \sigma_2), X_{b,i} \sim N(\mu, \psi_2). \tag{80}$$

The above model can be estimated in Mplus 8.1 using latent centering within the residual dynamic structural equation modeling (RDSEM) framework, see Asparouhov et al. (2018) and Asparouhov and Muthén (2018). The difference between this model and the standard two-level regression model (47–49) is described in equations (78–79), where we now allow the variables to be autocorrelated across time through their within level residuals via an autoregressive lag 1 model, also referred to as an AR(1) model. Because of equation (79) which allows for autocorrelation modeling for the predictor, this model is also different from the time-series models described in Chapter 6 in Raudenbush and Bryk (2002). Such time-series models are typically estimated with the observed centering and the REML estimator, see Bolger and Laurenceau (2013). In the econometrics literature, such models are used for pooled time-series and cross-sectional data, see Mundlak (1978) and Wooldridge (2008).

We generate data according to the above model with $N = 200$ individuals and varying number of time points T. The residual variance parameters are set to 1 and the intercept and mean parameters are set to 0. The parameter β_1 is set to 1, while the between level parameter β_2 takes values of -1, to create contextual effect, and 1 to generate data without contextual effect. The values of r_x and r_y are set to 0.7 or 0 to evaluate the autocorrelation effect on the estimation. With each set of parameters we generate 100 samples. The data will be analyzed with four different methods. The first method is based on the latent centering where the autocorrelations are taken into account, based on the RDSEM framework. The second method is using latent centering under the multilevel SEM framework where the autocorrelations are not modeled. The third method is based on the observed centering under the multilevel framework where the autocorrelations are not modeled. The fourth method uses the REML approach where the observed centering is used, the autocorrelation r_y is modeled and the autocorrelation r_x is ignored. This model will be estimated also within the RDSEM framework, using Bayesian estimation. It is well-known that both Bayesian estimation and REML estimation are asymptotically equivalent to ML estimation and therefore to each other. We verified that within individual replications the REML

TABLE 11
Absolute Bias(Coverage) for β_1

Time Points	*Contextual Effect*	r_y/r_x	*Latent Centering with autocorrelation*	*Latent Centering without autocorrelation*	*Observed Centering*	*REML Observed Centering with r_y*
30	Yes	0.7/0.7	0.00 (0.93)	0.00 (0.69)	0.00 (0.69)	0.00 (0.95)
60	Yes	0.7/0.7	0.00 (0.94)	0.00 (0.82)	0.00 (0.82)	0.00 (0.96)
100	Yes	0.7/0.7	0.00 (0.90)	0.00 (0.70)	0.00 (0.70)	0.00 (0.90)
30	No	0.7/0.7	0.00 (0.95)	0.00 (0.69)	0.00 (0.69)	0.00 (0.96)
100	No	0.7/0.7	0.00 (0.91)	0.00 (0.70)	0.00 (0.70)	0.00 (0.90)
30	Yes	0.7/0.0	0.00 (0.93)	0.00 (0.97)	0.00 (0.97)	0.00 (0.92)
30	Yes	0.0/0.7	0.00 (0.89)	0.00 (0.90)	0.00 (0.90)	0.00 (0.89)
30	Yes	0.0/0.0	0.00 (0.94)	0.00 (0.95)	0.00 (0.95)	0.00 (0.95)

TABLE 12
Absolute Bias (Coverage) for β_2

Time Points	*Contextual Effect*	r_y/r_x	*Latent Centering with autocorrelation*	*Latent Centering without autocorrelation*	*Observed Centering*	*REML Observed Centering with r_y*	*Analytically Derived Observed Centering Bias*
30	Yes	0.7/0.7	0.14(0.83)	0.44(0.01)	0.50(0.00)	0.51(0.00)	0.51
60	Yes	0.7/0.7	0.02(0.92)	0.26(0.18)	0.31(0.03)	0.30(0.05)	0.31
100	Yes	0.7/0.7	0.02(0.93)	0.16(0.54)	0.19(0.36)	0.18(0.37)	0.20
30	No	0.7/0.7	0.01(0.96)	0.00(0.98)	0.00(0.96)	0.00(0.96)	0.00
100	No	0.7/0.7	0.00(0.96)	0.00(0.95)	.00(0.95)	0.00(0.95)	0.00
30	Yes	0.7/0.0	0.00(0.93)	0.00(0.97)	0.07(0.89)	0.07(0.88)	0.13
30	Yes	0.0/0.7	0.16(0.85)	0.44(0.00)	0.50(0.00)	0.50(0.00)	0.51
30	Yes	0.0/0.0	0.01(0.98)	0.01(0.97)	0.07(0.84)	0.07(0.87)	0.13

results obtained from SAS, HLM and SPSS software packages are nearly identical to the RDSEM estimation for the fourth method. The simulation results are easier to obtain using the Mplus RDSEM estimation so here we report those results. For each set of parameters we generate 100 samples and analyze them with the four methods. The results for β_1 are reported in Table 11 and the results for β_2 are reported in Table 12.

The results in Table 11 show that all four estimation methods produce no bias in the parameter estimates for β_1. However, the observed and the latent centering methods based on the two-level SEM estimation, ignoring the autocorrelation, produce bias in the standard errors. This bias results in the lower coverage for β_1. The bias can be resolved by the REML method or by the latent centering method with autocorrelation modeling. This bias occurs when both r_y>0 and r_x>0 and disappears when either of the two autocorrelations are zero. The bias exists even with large samples, when there is a large number of time points, and with or without contextual effect. The size of this bias will depend on the size of the autocorrelations. The bias occurs due to overestimation of the number of independent observations on the within level when the autocorrelations are ignored. It is interesting to note here that the REML method which ignores the r_x autocorrelation is still able to resolve the bias in the standard error for β_1.

The results in Table 12 are more dramatic and show large parameter estimate bias for β_2 when r_x>0 and there is a contextual effect. The bias occurs for the observed and the latent centering methods without the autocorrelations, but also occurs for the REML estimation method, which includes r_y. This bias also results in lower coverage for the β_2 parameter. The size of the bias is affected by three components: the autocorrelation r_x, the size of the contextual effect and the number of time points T. The bias increases as r_x or the contextual effect increase and decreases as the number of time points increases. We can conclude from these results that the bias will disappear if r_x is small, the contextual effect is small or the number of time points is large. It appears, however, that a much larger number of time points maybe needed in time-series context (much larger than 100) to resolve this bias as compared to the size of the clusters needed to resolve Lüdtke's bias in regular multilevel regression. This is because the size of the bias is larger. The last row in Table 12 shows that Lüdtke's bias is 0.07 in the case when $T = 30$, while the bias of β_2 in the first row in the presence of the autocorrelations is 0.50, i.e., 7 times larger. It is interesting to note here that the bias due to the observed centering (Lüdtke's bias) is in the same direction as the bias due to ignoring r_x and so it adds up. The bias of the observed centering ignoring the autocorrelation and the bias in the REML method is essentially the sum of the bias of the latent centering method ignoring the autocorrelation r_x and Lüdtke's bias. Thus, both the REML method and the two-level method with observed centering show the largest bias for β_2. It is also interesting to note that based on the results in the last row in Table 12, we can see that the REML method is unable to resolve Lüdtke's bias as well.

The source of the bias for β_2 for time-series model is the same as that for Lüdtke's bias, i.e., the bias is due to not properly accounting for the measurement error in the mean of the covariate, which is used as a predictor on the between level. That measurement error is somewhat more complex in time-series settings as it depends not only on the number of time points T but also on the autocorrelation r_x. The results in Table 12 also point out that it is a common misconception to focus on modeling the autocorrelation for the dependent variable Y but to ignore the autocorrelation of the covariate X. Clearly, the value of r_x has a much larger impact on the estimation method than r_y. Ignoring r_y has consequences for the standard errors only, while ignoring r_x has larger consequences for the parameters estimates. This is important in light of the popularity of the REML method and its widespread use and availability in many statistical packages. This simulation shows that latent centering with autocorrelation modeling is the best method for time-series regressions.

Using the same approach as in Lüdtke et al. (2008), we can estimate the bias in the parameter estimates of β_2 for the estimation methods based on the observed centering in the context of time-series models. The analytic estimate for this bias is:

$$(\beta_1 - \beta_2)\frac{(1-ICC)/T^*}{ICC + (1-ICC)/T^*} \tag{81}$$

where

$$T^* = T\frac{1-r_x}{(1+r_x)(1-2r_x/(T(1-r_x^2)))}, \tag{82}$$

ICC is the intra-class correlation of X and T is the number of time points, see Appendix A. Thus, the bias in the time-series context can be computed the same way as Lüdtke's bias in the two-level context but now the cluster size T is reduced due to the autocorrelation r_x. The last column in Table 12 shows that the bias estimate is fairly accurate and that it matches the results obtained in the simulation study.

The bias in the parameter estimates of β_2 given in formula (81) can also be viewed as the bias of the contextual effect $\beta_2 - \beta_1$. Therefore, the relative bias for the contextual effect (with negative sign) is simply

$$\frac{(1-ICC)/T^*}{ICC + (1-ICC)/T^*}. \tag{83}$$

Note that this relative bias is independent of the size of the contextual effect. It only depends on the quantities T, r_x and ICC. Note also that the relative bias of the contextual effect is always between 0 and 1 and, thus, can be interpreted as the percentage of underestimation of the contextual effect. Figure 1 shows the relative bias as a function of ICC when $r_x = 0.5$ and $T = 30$. Figure 2 shows the relative bias as a function r_x when $ICC = 1/3$ and $T = 30$. Figure 3 shows the relative bias as a function T when $ICC = 1/3$ and $r_x = 0.5$. These figures indicate that small ICC values have the biggest impact on the relative bias, followed by large r_x values, followed by small T values.

Lagged centering: Nickell's bias

Let Y_{it} be the observed variable for individual i at time t. Consider the following two-level autoregressive model:

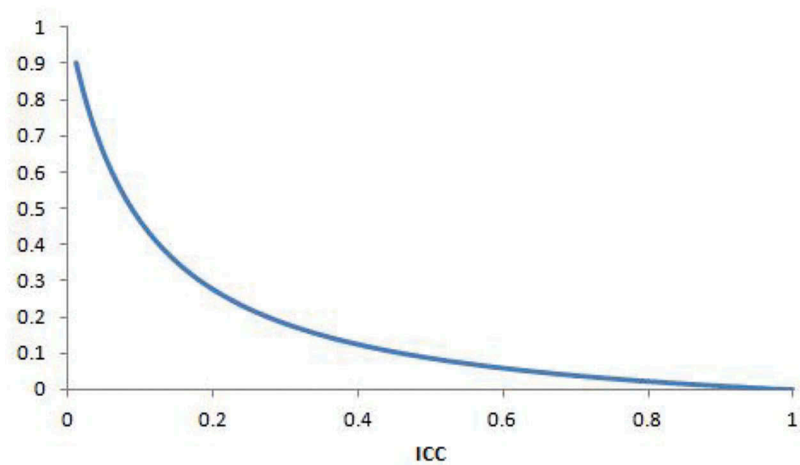

FIGURE 1 Relative contextual bias as a function of ICC, $r_x = 0.5$, $T = 30$.

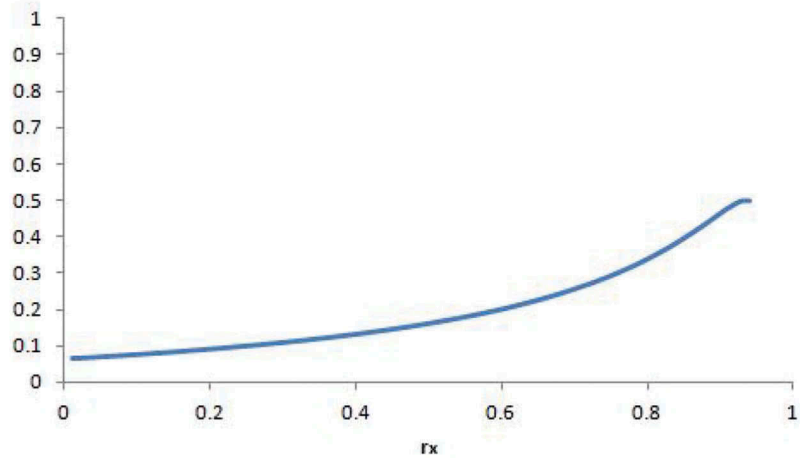

FIGURE 2 Relative contextual bias as a function of r_x, $ICC = 1/3$, $T = 30$.

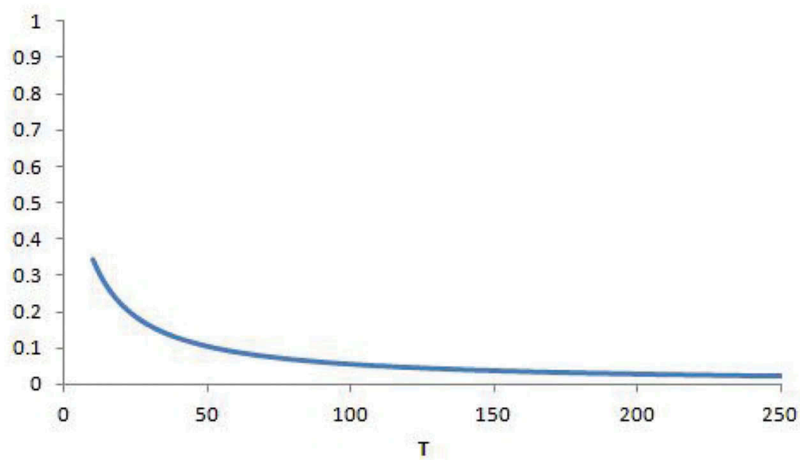

FIGURE 3 Relative contextual bias as a function of T, $ICC = 1/3$, $r_x = 0.5$.

$$Y_{it} - \mu_i = \phi_i(Y_{i,t-1} - \mu_i) + \xi_{it} \tag{84}$$

$$\mu_i = \mu + \varepsilon_{i1} \tag{85}$$

$$\phi_i = \phi + \varepsilon_{i2} \tag{86}$$

$$\xi_{it} \sim N(0,\sigma), \varepsilon_{i1} \sim N(0,\theta_1), \varepsilon_{i2} \sim N(0,\theta_2). \tag{87}$$

If we use the observed centering for the covariate $Y_{i,t-1}$ and we estimate the resulting two-level random slope regression with the ML estimator, the estimate of ϕ is biased. This bias is know as Nickell's bias and is approximately

$$-\frac{1+\phi}{T-1}, \tag{88}$$

where T is the number of observations in the time-series for each individual. The bias can be eliminated by using the latent centering for the covariate and estimating the model with the Bayes estimator in Mplus. Simulation results for the above model are presented in Asparouhov et al. (2018).

Lagged centering of categorical variables and the random tetrachoric autocorrelations

Let us consider now the corresponding autoregressive model for a categorical variable. For simplicity we will use a binary variable. The autoregressive model is expressed in terms of the underlying continuous variable Y^*_{it}

$$Y^*_{it} - \mu_i = \phi_i(Y^*_{i,t-1} - \mu_i) + \xi_{it} \tag{89}$$

$$\mu_i = \mu + \varepsilon_{1i} \tag{90}$$

$$\phi_i = \phi + \varepsilon_{2i} \tag{91}$$

$$P(Y_{it} = 1) = P(Y_{it}^* > 0) \tag{92}$$

$$\xi_{it} \sim N(0,1), \varepsilon_{i1} \sim N(0,\theta_1), \varepsilon_{i2} \sim N(0,\theta_2). \tag{93}$$

The above model is essentially a multilevel time-series model for an observed binary variable. Such a model is a powerful alternative to Markov chain modeling as it can be incorporated within a general multilevel model and it allows for cluster specific transition probabilities through the random tetrachoric autocorrelation ϕ_i. The model easily extends to ordered categorical variables as well as models with covariates and latent variables. In addition, the model can accommodate lagged relationships for lags greater than 1. Markov chain models by definition do not allow lagged relations beyond lag 1. Furthermore, the above model can easily accommodate multivariate time-series modeling with categorical variables which is a valuable alternative to the latent Markov chain modeling, see Asparouhov, Hamaker, and Muthén (2017).

Note that this model is substantially different than the categorical time-series models discussed in Asparouhov et al. (2018). The difference is in the fact that the autocorrelation in the above model is used directly for the underlying Y_{it}^* rather than a latent variable measured by Y_{it}^*. This difference is important and results in a substantially faster estimation algorithm and can be used with much shorter time-series data, i.e., the model can be estimated with as few as 20 observations per person.

In this section, we explore the possible alternatives and approximations to the above model. Our goal is to determine if Nickell's bias can be observed in these settings as well. We generate data according to the above model using the following parameter values $\mu = 0$, $\phi = 0.3$, $\theta_1 = 0.6$ and $\theta_2 = 0.01$. To evaluate the bias we generate a single data set with 5000 clusters. Such a data set represents a large sample that can be used to determine the bias of the estimates. This approach can be used as an alternative to the usual approach of using multiple smaller data sets. In the simulation, we need to standardize the random effects in each cluster separately and this is easily done in Mplus with one large data set rather than multiple smaller data sets. We generate the data using cluster size 20, 50 and 200.

In the continuous case, when the observed centering is used and there is no missing data, the time-series model is reduced to a standard two-level regression. There are no such easy simplifications for the model in (89–93). We analyze the data using five different methods. The first method is the latent centering option where we estimate (89–93) using the Bayes estimator in Mplus. The second method is based on the observed centering for the predictor using the observed scale as suggested by Enders and Tofighi (2007). We used that observed centering approach in the previous section as well. The third method is using the uncentered observed predictor. In the continuous case the uncentered method performed well, see Hamaker and Grasman (2015), and, therefore, it is worthwhile to explore this method as well for the categorical case. The fourth and the fifth methods are both using the predictor on the latent scale. These methods simplify the model by assuming that the predictor is a separate variable rather than the lagged copy of the dependent variable. Such simplification is often used in the continuous case and it has been used to estimate multivariate time-series models from variance covariance matrices of the dependent variables and its lagged variables, see Zhang, Hamaker, and Nesselroade (2008). Using this approach we can estimate the above model by estimating a bivariate multilevel probit model. The fourth method uses latent centering for the covariate while the fifth method does not, but are both based on this bivariate probit modeling. Note, however, that this bivariate modeling simplification does not lead to substantial simplifications in the estimation of the models. Both methods can be estimated only in Mplus 8.1.

With all five estimation methods, we standardize the within level regression coefficient within each cluster so that it represents a correlation parameter and we average that across the clusters. This is accomplished in Mplus with the option OUTPUT: STAND(CLUSTER). The results of the simulation study are presented in Table 13. Clearly the

TABLE 14
Bias and Coverage for the Mean of the Random Tetrachoric Autocorrelation, $\phi = 0.3$

Cluster Size	*Latent Centering*
20	0.01 (0.98)
50	0.01 (0.90)
200	0.00 (0.98)

TABLE 13
Nickell's Bias for the Random Tetrachoric Autocorrelation, $\phi = 0.3$

Cluster Size	*Latent Centering*	*Observed Centering*	*Uncentered*	*Bivariate Centered*	*Bivariate Uncentered*
20	0.01	−0.16	−0.05	−0.08	0.04
50	0.00	−0.12	−0.06	−0.03	0.03
200	0.00	−0.09	−0.05	0.01	0.03

latent centering is the only acceptable alternative. The estimates are unbiased even with small cluster sizes. Among the rest of the methods the bivariate centered method appears to be the only one that could be of interest because the bias disappears as the cluster size increases. That method, however, does not provide proper standard error estimation. It is also clear from these results that Nickell's bias is far more complex in the categorical case and clearly the bias does not disappear for large cluster sizes.

We conduct one additional simulation study to evaluate the quality of the latent centering estimation over multiple replications. Using the same settings as above, we generate 100 samples with 100 clusters each and analyze the data using the latent centering. The results are presented in Table 14. The latent centering shows no bias and the coverage is near the nominal levels.

TWO-LEVEL MEDIATION

In this section, we consider the impact of centering in multilevel mediation modeling with random slopes. We use the 2-1-1 case for illustration purposes, see Preacher et al. (2010). Suppose that the dependent variable is Y_{ij}, the mediator variable is M_{ij} and the between level predictor variable is X_j. Than, the mediation model is as follows:

$$Y_{ij} = Y_{w,ij} + Y_{b,j} \tag{94}$$

$$M_{ij} = M_{w,ij} + M_{b,j} \tag{95}$$

$$Y_{w,ij} = \beta_{1,j} M_{w,ij} + \varepsilon_{w,ij} \tag{96}$$

$$Y_{b,j} = \alpha_1 + \beta_2 M_{b,j} + \beta_3 X_j + \varepsilon_{b,j}. \tag{97}$$

$$M_{b,j} = \alpha_2 + \beta_4 X_j + \xi_{b,j} \tag{98}$$

$$\varepsilon_{w,ij} \sim N(0, \sigma_w), M_{w,ij} \sim N(0, \psi_w) \tag{99}$$

$$\varepsilon_{b,j} \sim N(0, \sigma_b), \beta_{1,j} \sim N(\beta_1, \theta), \xi_{b,j} \sim N(0, \psi_b). \tag{100}$$

If we estimate the above model using the latent or the observed group mean centering methods for the mediator M_{ij}, the indirect effect from X to Y can be computed as $\beta_2\beta_4$. In Preacher et al. (2010), Appendix F, the hybrid method is used to estimate the model and in that case the indirect effect is computed as $(\beta_1 + \beta_2)\beta_4$. McNeish (2017) considers this model without the random slope and used the observed group mean centering method. We conduct a simulation study to compare these three centering methods. We generate 100 samples with C clusters of size L using the above model with the following parameters: $\sigma_w = 1$, $\psi_w = 1$, $\sigma_b = .9$, $\psi_b = .9$, $\theta = 1$, $\beta_2 = 1$, $\beta_3 = 0.5$, $\beta_4 = 1$, $\alpha_1 = 2$ and $\alpha_2 = 1$. The between level covariate X_j is generated from a standard normal distribution. The parameter β_1 takes two possible values in the simulations: −0.5 and 1. The first value corresponds to non-zero contextual effect, while the second value corresponds to zero contextual effect. We use three values for C : 500, 20 and 15. The first corresponds to a large number of clusters and the second and third to a small number of clusters. We also use three values for L : 50, 20 and 15.

The results are presented in Table 15. The hybrid method appears to be unbiased when the number of clusters is large but it seems to lose that advantage when the number of clusters is small. Regardless of that, however, the hybrid method has a substantially larger MSE as compared to the latent centering method and thus it can not be recommended. The observed group mean centering method works well when the contextual effect is small or when the size of the clusters is large. For smaller cluster sizes when the contextual effect is not 0, the indirect effect estimate is biased and the coverage drops below the nominal level. The latent centering method appears to work well in all cases, however, when the contextual effect is 0 and the number of clusters is small, the observed group mean centering method appears to have a slight MSE advantage. McNeish (2017) appears to have reached a similar conclusion for the two-level mediation model without random slopes, but fails to notice that this advantage clearly applies only to the case of no contextual effect. Table 17 shows that in the presence of a contextual effect and a small number of

TABLE 15
Indirect Effect: Absolute bias/coverage/SMSE

Number of Clusters	*Cluster Size*	*Contextual effect*	*Latent*	*Observed*	*Hybrid*
500	50	No	0.00/0.94/0.064	0.00/0.92/0.063	0.01/0.93/0.122
500	50	Yes	0.00/0.93/0.066	0.03/0.90/0.071	0.01/0.92/0.122
500	20	No	0.01/0.93/0.058	0.01/0.94/0.057	0.00/0.98/0.106
500	20	Yes	0.01/0.95/0.062	0.09/0.77/0.104	0.00/0.97/0.106
20	20	No	0.06/0.89/0.376	0.03/0.88/0.357	0.05/0.88/0.559
20	20	Yes	0.07/0.93/0.393	0.56/0.65/0.856	0.52/0.71/0.637
15	15	No	0.03/0.98/0.528	0.01/0.91/0.423	0.04/.90/.746
15	15	Yes	0.08/0.97/0.486	0.58/0.68/0.901	0.49/.67/.692

TABLE 16
Replicating MCNEISH (2017), TABLE 3: Indirect Effect Coverage for the Latent Centering ML-SEM Method

Number of Clusters	*Small Clusters Small effect*	*Large Clusters Small effect*	*Small Clusters Medium effect*	*Large Clusters Medium effect*
10	0.98	0.96	0.95	0.88
15	0.97	0.93	0.93	0.87
25	0.95	0.93	0.92	0.90
50	0.93	0.95	0.93	0.94
100	0.93	0.95	0.94	0.94

TABLE 17
Indirect Effect for Correlated Random Slope and Between Mediator

Centering	*Latent*	*Hybrid-Appendix F Preacher et al. (2010)*	*Hybrid*
Formula	$\beta_2\beta_4$	$(\beta_1+\beta_2)\beta_4$	$(\beta_1+\beta_2+\alpha_2\rho/\psi_b)\beta_4$
Bias	0.00	−0.77	0.00
Coverage	0.93	0.00	0.90
SMSE	0.057	0.782	0.149

clusters, the observed group mean centering is dramatically worse than both the latent centering and the hybrid method. While in the above simulation, we did not vary the size of the contextual effect, the size actually matters and the bigger the contextual effect the bigger its impact is on the bias of the observed and the hybrid method.

The main argument in McNeish (2017) against using latent centering with small number of clusters is the fact that the standard errors for the indirect effects are biased. Using partial information we obtained from the author regarding this finding we replicated the simulation study and found no such bias. The results we obtained are reported in Table 16 and clearly contradict the results in Table 3 of McNeish (2017), under ML-SEM-Delta method. The results in Table 16 show coverage near the nominal level, while the results in Table 3 of McNeish (2017) show coverage values as low as 0.63. In this simulation study, we used 500 replications and varying number of clusters. The cluster sizes are set to 10 to generate data with small clusters and to 100 to generate data with large clusters. As in McNeish (2017), we used a non-random slope on the within level, i.e., the variance θ of $\beta_{1,j}$ is set to 0 and $\beta_{1,j}$ is identical to the non-random slope β_1. The data is generated without contextual effect and $\beta_1=\beta_2=0.5$. To generate data with small effect size we set $\beta_3=0.3$ and $\beta_4=0.2$. To generate data with medium effect size we set $\beta_3=0.6$ and $\beta_4=0.4$. The remaining parameters in model (94–100) are set as follows: $\alpha_1=\alpha_2=0$, $\sigma_w=\psi_w=1$, $\sigma_b=\psi_b=.25$ and $X_j\sim N(0,.25)$. The results in Table 16 are obtained using the delta method for computing the standard errors of the indirect effect which is the Mplus default option.

We summarize our findings as follows. We do not recommend using the hybrid method at all. That is in contrast to Appendix F in Preacher et al. (2010). Better methods are now available for estimating multilevel mediation models with random slopes. When the size of the clusters is 100 or more, the latent centering method and the observed group mean centering method are essentially equivalent and their performance would be the same in terms of MSE, bias and coverage. In that case preference should be given to convenience, for example, if the ML estimator is desired, the observed group mean centering should be used because the latent centering is not available with the ML estimator. When the cluster sizes are smaller than 100, we recommend the latent centering method, except possibly when the number of clusters is smaller than 20 and the contextual effect is small as well. In that special case, the observed group mean centering method has an advantage in terms of MSE. However, since that advantage is somewhat minor and is conditional on the absence of contextual effect, which can be evaluated only after the estimation is completed, it is difficult to make a general recommendation for this method. The latent centering method can be considered the most universally applicable method.

Next we turn our attention to a slight modification of the above model that has somewhat surprising consequences for the hybrid and the uncentered method. Given that the hybrid method has been used in Preacher et al. (2010), it is worth to point out these complications. Consider the situation when the random slope is correlated with the between level component of the mediator variable.

$$\begin{pmatrix}\beta_{1,j}\\ \xi_{b,j}\end{pmatrix}\sim N\left(\begin{pmatrix}\beta_1\\ 0\end{pmatrix},\begin{pmatrix}\theta & \rho\\ \rho & \psi_b\end{pmatrix}\right). \quad (101)$$

Thus, we have added the parameter ρ to the model (94–100). It turns out that under these circumstances the formula $(\beta_1+\beta_2)\beta_4$ is no longer a valid way to compute the indirect effect. Using the methodology described in Muthén and Asparouhov (2015) and chapters 4 and 8 in Muthén et al. (2017), we compute the indirect effect as:

$$(\beta_1+\beta_2+\alpha_2\rho/\psi_b)\beta_4, \quad (102)$$

see Appendix B, which clearly differs from the formula used in Appendix F in Preacher et al. (2010).

We now conduct a simulation study to verify that this new formula (120) is a more accurate way for computing the indirect effect. We generate 100 samples with 500 clusters of size 50 using the above model with the contextual effect and the new parameter $\rho = 0.7$. Table 17 contains the results of the simulation study. We can make several conclusions from these results. The simulation confirms that the formula $(\beta_1 + \beta_2)\beta_4$ used in Preacher et al. (2010) does not compute the indirect effect correctly in these circumstance while formula (120) works fine in terms of being unbiased and the coverage is near the nominal level. Note also that the latent centering again outperformed the hybrid method in terms of being more accurate (smaller MSE). This example also illustrates another advantage of the latent centering method. Because the latent centering method completely separates the within and the between effects, the indirect effect can be computed always using the same simple formula $\beta_2\beta_4$ that is used also for computing indirect effects in single level models. On the other hand, the hybrid method computation of the indirect effect is more complex and likely to depend on the particular model. In addition, if the hybrid model is more complex there is no guarantee that the estimated model would be a good model to base our inference on. If the true model is the latent centering model, using uncentered method during the estimation could lead to other biases that can lead to further problems with the estimation of the indirect effect. For example, if the latent centering model is misspecified and the parameter ρ is fixed to 0, there is no big change in the estimation of the indirect effect and the estimate is still unbiased. If that parameter is fixed to 0 with the hybrid model, the estimation of the indirect effect will be biased and no matter what formula is used the indirect effect will be biased because the model is misspecified. Thus, we conclude that yet another advantage of the latent centering method for multilevel mediation models with random slopes is the fact that the computation of the indirect and the direct effects is more robust to model misspecifications.

The supplemental materials of Preacher et al. (2010) will be updated to reflect the above finding. Other appendices in Preacher et al. (2010) that may require additional development are appendices G, J and O, however, such development may be too complex and the latent centering approach may remain the only practical method for conducting mediation models with random slopes. All other supplemental materials in Preacher et al. (2010) are correct as they do not involve random slopes.

CENTERING OF PREDICTOR OR MEDIATOR VARIABLE WITH MISSING DATA

In this section, we consider the different centering options when the predictor or mediator variable has missing data. Observed group mean centering is based on the sample cluster means. The sample mean, however, is a biased estimate of the mean when the missing data is MAR. It is unbiased only when the missing data is MCAR. Thus, we can expect that the observed group mean centering will perform poorly in the presence of MAR missing data. We conduct a simulation study to illustrate this point based on the mediation model described in the previous section. We use the model parameters corresponding to the first row in Table 15, that is, we use the model without contextual effect, with 500 clusters and cluster size of 50. Under these circumstance and no missing data the observed group mean centering performs well. Two changes are made to the parameters, however. For data generation purposes the two mean parameters α_1 and α_2 are set to 0. This is done to simplify the understanding of missing data generation. To generate MAR missing data for the mediator we use the following model:

$$P(M_{ij} \text{ is missing}) = \frac{1}{1 + Exp(1 + 0.5Y_{ij})}. \tag{103}$$

This model generation produces approximately 33% missing data. To generate MCAR missing data for the mediator we use the model:

$$P(M_{ij} \text{ is missing}) = \frac{1}{1 + Exp(1)} \tag{104}$$

which produces approximately 27% missing data. We use four different estimation methods: the latent centering method using Bayesian estimation, the observed group mean centering method using Bayesian estimation, the observed group mean centering method using ML estimation in Mplus with montecarlo numerical integration and the observed group mean centering method using ML estima-

TABLE 18
Indirect Effect with Missing Data on the Mediator: Absolute bias/coverage/SMSE

Missing Data	*Latent Bayes*	*Observed Bayes*	*Observed ML + montecarlo*	*Observed ML + listwise*
MCAR	0.00/0.93/0.064	0.01/0.92/0.062	0.02/0.91/0.068	0.00/0.91/0.064
MAR	0.00/0.92/0.063	0.09/0.68/0.108	0.10/0.58/0.121	0.13/0.39/0.142
Comp Time	3 sec	5 sec	16 min	1 sec

tion and listwise deletion of the missing data based on estimation without numerical integration. We evaluate the performance of the four estimation methods by computing the absolute bias, the coverage and the MSE for the indirect effect. The results of the simulation are presented in Table 18.

The results clearly show that the observed group mean centering method, with any of the three estimators, yields biased results when the missing data is MAR. The coverage is also affected and is below the nominal levels. The latent centering method on the other hand performs well and the coverage is near the nominal level. The SMSE differences are substantial among the four estimators. The latent centering method is the most accurate followed by the observed group mean centering method using the Bayesian estimator. The observed group mean centering method using three dimensional montecarlo integration is less accurate than the Bayes estimator due to the imprecision of the montecarlo integration. We use the Mplus default setting for the numerical integration. If the number of integration points is increased further the precision of the estimates should become closer to the ones obtained with the observed group mean centering method and the Bayesian estimator. However, Table 18 also shows the computational times for each replication for the four estimation methods. All methods are fast except the ML method with montecarlo integration. Thus, increasing the number of integration points to improve precision is impractical. The most inaccurate method is the listwise deletion method, in terms of bias, MSE and coverage. When the missing data is MCAR all methods perform similarly: no bias, good coverage and similar MSE. This further shows that the biased results seen in Table 18 for the MAR missing data and the estimators with observed group mean centering is precisely due to the fact that the data is not MCAR. Note also that unlike Lüdtke's bias or the indirect effect bias in the presence of contextual effect the bias due to MAR missing data will not disappear if the size of the clusters is increased. That is due to the fact that, in the case of MAR missing data, the error in the sample mean estimate of the true mean will not disappear as sample size increases.

SUMMARY

In this note, we show that the latent centering of predictors and mediators in multilevel models can be used to resolve multiple problems that occur with the traditional observed centering. Among these are Lüdtke's bias in the estimation of contextual effects, Nickell's bias in the estimation of the autocorrelations, bias due to MAR missing data, bias due to the non-linearity of the link function when the predictor is a categorical variable, and bias in the indirect effect estimate in multilevel mediation models. The method also provides a clean separation of the within and the between effects which makes the multilevel mediation models much easier to utilize. In most situations, the benefits extend beyond the elimination of the bias and also result is much more accurate estimates as measured by the MSE as well as more accurate standard errors. The use of the Bayesian estimation makes it possible to go beyond traditional multilevel methodology. The proposed latent variable centering gives unexpected improvements over observed variable centering.

In almost all of the simulations presented in this article, the latent centering method based on the Bayesian estimation outperformed the alternatives. The one notable exception is the second to last row in Table 15, which is characterized by small cluster sizes, small number of clusters and no contextual effect. Under such circumstances the observed centering is more accurate and yields smaller MSE. In the context of non-random slope regression and the ML estimator, Lüdtke et al. (2008) also find that the observed centering can outperform the latent centering in certain data constellations (e.g., small ICC of the covariate, small clusters sizes and small number of clusters). Asymptotically, as the number of level 2 units increases, the latent centering method is guaranteed to outperform the observed centering method. The small sample size results, however, should not be ignored as many practical applications are based on such small samples. Further small sample size research is needed to identify the circumstances where the observed centering method is the more accurate approach. In addition, in small sample size estimation, the ML estimator is not necessarily similar to the Bayes estimator and the pairing of estimator and centering option should be investigated as well. Furthermore, the choice of the priors in the Bayesian estimation with small sample size becomes important and should be evaluated as well. Note, however, that small sample size results are extremely difficult to generalize and to extrapolate from one situation to another. Any claims on estimation performance specific to small sample size should not be relied upon unconditionally.

ACKNOWLEDGMENTS

We thank Ellen Hamaker, Kristopher Preacher, Michael Zyphur, David MacKinnon, Craig Enders, Jean-Philippe Laurenceau, and Herbert Marsh for helpful discussion and comments on the earlier draft of this paper.

REFERENCES

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2017). Dynamic latent class analysis. *Structural Equation Modeling*, *24*, 257-269. doi:10.1080/10705511.2016.1253479

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling, 25*, 359–388. doi:10.1080/10705511.2017.1406803

Asparouhov, T., & Muthén, B., (2006a). Constructing covariates in multilevel regression, Mplus Web Notes 11, Retrieved from http://statmodel.com/download/webnotes/webnote11.pdf

Asparouhov, T., & Muthén, B., (2006b). Multilevel modeling of complex survey data. Proceedings of the Joint Statistical Meeting in Seattle, August 2006. ASA Section on Survey Research Methods, 2718–2726.
Asparouhov, T., & Muthén, B., (2010) Bayesian analysis using Mplus: Technical implementation. Retrieved from http://statmodel.com/download/Bayes3.pdf
Asparouhov, T., & Muthén, B. (2018). Comparison of RDSEM and DSEM. Technical report. Retrieved from http://statmodel.com/
Bolger, N., & Laurenceau, J. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York, NY: Guilford Press.
Enders, C., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121–138. doi:10.1037/1082-989X.12.2.121
Hamaker, E. L., & Grasman, R. P. P. P. (2015). To center or not to center? Investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, *5*, 1492. doi:10.3389/fpsyg.2014.01492
Lai, M. H. C., & Kwok, O. (2015). Examining the rule of thumb of not using multilevel modeling: The design effect smaller than two rule. *The Journal of Experimental Education*, *83*(3), 423–438. doi:10.1080/00220973.2014.907229
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. doi:10.1037/a0012869
Marsh, H. W., Kuyper, H., Morin, A., Parker, P., & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, *33*, 50–66. doi:10.1016/j.learninstruc.2014.04.002
Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A., Abduljabbar, A., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47*, 106–124. doi:10.1080/00461520.2012.670488
Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*, 764–802. doi:10.1080/00273170903333665
McNeish, D. (2017). Multilevel mediation with small samples: A cautionary note on the multilevel structural equation modeling framework, *Structural Equation Modeling*,*24*, 609–625. doi:10.1080/10705511.2017.1280797
Mundlak, Y. (1978). Pooling of Time-series and cross-section data. *Econometrica: Journal of the Econometric Society*, *46*, 69–85. doi:10.2307/1913646
Muthén, B. (1990). Moments of the censored and truncated bivariate normal distribution. *British Journal of Mathematical and Statistical Psychology*, *43*, 131–143. doi:10.1111/bmsp.1990.43.issue-1
Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*, 376–398. doi:10.1177/0049124194022003006
Muthén, B., & Asparouhov, T. (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling*, *22*, 12–23. doi:10.1080/10705511.2014.935843
Muthén, B., Muthén, L., & Asparouhov, T. (2017). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén & Muthén.
Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, *25*, 267–316. doi:10.2307/271070
Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, *49*(6), 1417–1426.
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. doi:10.1037/met0000052
Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*, 209–233. doi:10.1037/a0020141
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage Publications.
Verhagen, A. J., & Fox, J.-P. (2012). Bayesian tests of measurement invariance. *British Journal of Mathematical and Statistical Psychology, 66*(03), 383–401.
Wooldridge, J. (2008). *Introductory econometrics: A modern approach*. Mason, OH: South-Western Cengage Learning.
Zhang, Z., Hamaker, E., & Nesselroade, J. (2008). Comparisons of four methods for estimating a dynamic factor model. *Structural Equation Modeling*, *15*, 377–402. doi:10.1080/10705510802154281

APPENDIX A. LÜDTKE'S BIAS IN TIME-SERIES MODELS

According to formula (53), Lüdtke's bias is equal to

$$(\beta_1 - \beta_2)(1 - R) \tag{105}$$

where R is the reliability of the sample cluster mean for X_{ij}

$$R = \frac{Var(X_{b,j})}{Var(\overline{X}_{.j})}. \tag{106}$$

Denote by $\psi_w = Var(X_{w,ij})$ and by $\psi_b = Var(X_{b,j})$. Note that $\overline{X}_{.j} = X_{b,j} + \overline{X}_{w,.j}$. Thus $Var(\overline{X}_{.j}) = \psi_b + Var(\overline{X}_{w,.j})$. The variance $Var(\overline{X}_{w,.j})$ can be computed as follows

$$Var(\bar{X}_{w,.j}) = \frac{1}{T^2} Cov(X_{w,1j} + X_{w,2j} + \ldots + X_{w,Tj}, X_{w,1j} + X_{w,2j} + \ldots + X_{w,Tj}) = \tag{107}$$

$$\frac{1}{T^2}(T\psi_w + 2(T-1)r_x\psi_w + 2(T-2)r_x^2\psi_w + \ldots + 2r_x^{T-1}\psi_w) = \tag{108}$$

$$\frac{\psi_w}{T^2}(-T + 2\frac{1-r_x}{1-r_x} + 2\frac{1-r_x^2}{1-r_x} + \ldots + 2\frac{1-r_x^T}{1-r_x}) = \tag{109}$$

$$\frac{\psi_w}{T^2(1-r_x)}(-T(1-r_x) + 2T - 2(r_x + r_x^2 + \ldots + r_x^T)) = \tag{110}$$

$$\frac{\psi_w}{T^2(1-r_x)}(T(1+r_x) - \frac{2r_x(1-r_x^T)}{1-r_x}) = \tag{111}$$

$$\frac{\psi_w(1+r_x)}{T(1-r_x)}(1 - \frac{2r_x(1-r_x^T)}{T(1-r_x^2)}) \approx \tag{112}$$

$$\frac{\psi_w(1+r_x)}{T(1-r_x)}(1 - \frac{2r_x}{T(1-r_x^2)}) = \frac{\psi_w}{T^*}, \tag{113}$$

where in the above approximation we used the fact that $|r_x|<1$ and $r_x^T \approx 0$ for sufficiently large T. We also used the definition of T^* given in equation (82). Thus, Lüdtke's bias for the time-series model is

$$(\beta_1 - \beta_2)(1 - R) = (\beta_1 - \beta_2)(1 - \frac{\psi_b}{\psi_b + \psi_w/T^*}) = \tag{114}$$

$$(\beta_1 - \beta_2)\frac{\psi_w/T^*}{\psi_b + \psi_w/T^*} = (\beta_1 - \beta_2)\frac{(1 - ICC)/T^*}{ICC + (1 - ICC)/T^*}, \tag{115}$$

where ICC is the intra-class correlation of X.

APPENDIX B. COMPUTING THE INDIRECT EFFECT FOR THE 2–1-1 MEDIATION MODEL

When the 2–1-1 mediation model is estimated with the uncentered or the hybrid methods, the between part of Y_{ij} includes also $\beta_{1,j}M_{b,j}$ which is the term that appears on the within level because the mediator is not centered. Thus with the uncentered or the hybrid methods

$$Y_{b,j} = \alpha_1 + \beta_2 M_{b,j} + \beta_3 X_j + \beta_{1,j} M_{b,j} + \varepsilon_{b,j}. \tag{116}$$

If the random effects $M_{b,j}$ and $\beta_{1,j}$ are correlated as shown in equation (101), we can use equation (98) to get that the conditional expectation of $\beta_{1,j}$ conditional on $M_{b,j}$ and X_j

$$E[\beta_{1,j}|M_{b,j}, X_j] = \beta_1 + (M_{b,j} - \alpha_2 - \beta_4 X_j)\rho/\psi_b. \tag{117}$$

Combining equations (116) and (117) we get the conditional expectation for $Y_{b,j}$

$$\begin{aligned} E[Y_{b,j}|M_{b,j}, X_j] = \alpha_1 + \beta_2 M_{b,j} + \beta_3 X_j \\ + (\beta_1 + (M_{b,j} - \alpha_2 - \beta_4 X_j)\rho/\psi_b)M_{b,j} \end{aligned} \tag{118}$$

$$\begin{aligned} = \alpha_1 + (\beta_1 + \beta_2 - \alpha_2\rho/\psi_b)M_{b,j} + \beta_3 X_j \\ + (\rho/\psi_b)M_{b,j}^2 - \beta_4(\rho/\psi_b)M_{b,j}X_j. \end{aligned} \tag{119}$$

Using the potential outcome methodology described in Muthén and Asparouhov (2015) and chapters 4 and 8 in Muthén et al. (2017), we can now compute the total natural indirect effect (TNIE)

$$TNIE = E[Y(1, M(1))] - E[Y(1, M(0))]. \tag{120}$$

To compute $E[Y(1, M(1))]$ we use the above conditional expectation formula (119), where $M_{b,j}$ is now $M(1)$ and is distributed as $M(1) \sim [M_{b,j}|X_j = 1] = N(\alpha_2 + \beta_4, \psi_b)$

$$E[Y(1, M(1))] = E[Y_{b,j}|M_{b,j} \sim N(\alpha_2 + \beta_4, \psi_b), X_j = 1] = \tag{121}$$

$$\begin{aligned} \alpha_1 + (\beta_1 + \beta_2 - \alpha_2\rho/\psi_b)(\alpha_2 + \beta_4) + \beta_3 \\ + (\rho/\psi_b)((\alpha_2 + \beta_4)^2 + \psi_b) \\ - \beta_4(\rho/\psi_b)(\alpha_2 + \beta_4). \end{aligned} \tag{122}$$

To compute $E[Y(1, M(0))]$, we use the conditional expectation formula (119), where $M_{b,j}$ is $M(0)$ and is distributed as $M(0) \sim [M_{b,j}|X_j = 0] = N(\alpha_2, \psi_b)$

$$E[Y(1, M(0))] = E[Y_{b,j}|M_{b,j} \sim N(\alpha_2, \psi_b), X_j = 1] = \tag{123}$$

$$\begin{aligned} \alpha_1 + (\beta_1 + \beta_2 - \alpha_2\rho/\psi_b)\alpha_2 + \beta_3 \\ + (\rho/\psi_b)(\alpha_2^2 + \psi_b) - \beta_4(\rho/\psi_b)\alpha_2. \end{aligned} \tag{124}$$

Subtracting these two expressions yields

$$TNIE = (\beta_1 + \beta_2 + \alpha_2\rho/\psi_b)\beta_4. \tag{125}$$